Title of the invention

M. TUBERCULOSIS ANTIGENS

Related applications

This application is a continuation-in-part of:

US application Serial No. 09/289,388 filed 12 April 1999, which is a continuation of US application Serial No. 08/465,640 filed 5 June 1995, now US Patent No. 5,955,077, issued September 21, 1999, which is a continuation-in-part of US 08/123,182 filed 20 September 1993, now abandoned, and a continuation-in-part of PCT/DK94/00273, filed July 1, 1994, published as WO95/01441, and claiming priority from Danish application 0798/93, filed July 2, 1993;

US application Serial No. 09/050,739 filed 30 March 1998, which is claims priority from US provisional application Serial No. 60/044,624 filed 18 April 1997;

Andersen et al., application Serial No.09/791,171, filed 20 February 2001, as a divisional of U.S. application Serial No. 09/050,739; and

US application Serial No. 09/246,191, filed 30 December 1998, which claims priority from US provisional 60/070,488 filed 5 January 1998.

Reference is also made to commonly-owned U.S. Patent No. 6,120,776.

Each of these patents, patent applications and patent publications, as well as all documents cited in the text of this application, and references cited in the documents referred to in this application (including references cited in the aforementioned patents, patent applications and patent publications or during their prosecution) are hereby incorporated herein by reference.

Field of invention

The present application discloses new immunogenic polypeptides and new immunogenic compositions based on polypeptides derived from the short time culture filtrate of M. tuberculosis.

General Background

Human tuberculosis caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) is a severe global health problem, responsible for approx. 3 million deaths annually, according to the WHO. The worldwide incidence of new tuberculosis (TB) cases had been falling during the 1960s and 1970s but during recent years this trend has markedly changed in part due to the advent of AIDS and the appearance of multidrug resistant strains of *M. tuberculosis*.

The only vaccine presently available for clinical use is BCG, a vaccine whose efficacy remains a matter of controversy. BCG generally induces a high level of acquired resistance in animal models of TB, but several human trials in developing countries have failed to demonstrate significant protection. Notably, BCG is not approved by the FDA for use in the United States because BCG vaccination impairs the specificity of the Tuberculin skin test for diagnosis of TB infection.

This makes the development of a new and improved vaccine against TB an urgent matter, which has been given a very high priority by the WHO. Many attempts to define protective mycobacterial substances have been made, and different investigators have reported increased resistance after experimental vaccination. However, the demonstration of a specific long-term protective immune response with the potency of BCG has not yet been achieved.

Immunity to *M. tuberculosis* is characterized by some basic features; specifically sensitized T lymphocytes mediates protection, and the most important mediator molecule seems to be interferon gamma (IFN-γ).

M. tuberculosis holds, as well as secretes, several proteins of potential relevance for the generation of a new TB vaccine. For a number of years, a major effort has been put into the identification of new protective antigens for the development of a novel vaccine against TB. The search for candidate molecules has primarily focused on proteins released from dividing bacteria. Despite the characterization of a large number of such proteins only a few of these have been demonstrated to induce a protective immune response as subunit vaccines in animal models, most notably ESAT-6 and Ag85B (Brandt et al 2000 Infect. Imm. 68:2; 791-795).

In 1998 Cole et al published the complete genome sequence of *M. tuberculosis* and predicted the presence of approximately 4000 open reading frames (Cole et al 1998). Following the sequencing of the *M. tuberculosis* genome, nucleotide sequences comprising Rv0288, Rv3019c or Rv3017c are described in various databases and putative protein sequences for the above sequences are suggested, Rv3017c either comprising methionin or valine as the first amino acid (The Sanger Centre database (http://www.sanger.ac.uk/Projects/M_tuberculosis), Institut Pasteur database (http://genolist.pasteur.fr/TubercuList) and GenBank (http://www4.ncbi.nlm.nih.gov)). However important, this sequence information cannot be used

to predict if the DNA is translated and expressed as proteins *in vivo*. More importantly, it is not possible on the basis of the sequences, to predict whether a given sequence will encode an immunogenic or an inactive protein. The only way to determine if a protein is recognized by the immune system during or after an infection with *M. tuberculosis* is to produce the given protein and test it in an appropriate assay as described herein.

Diagnosing *M. tuberculosis* infection in its earliest stage is important for effective treatment of the disease. Current diagnostic assays to determine *M. tuberculosis* infection are expensive and labour-intensive. In the industrialised part of the world the majority of patients exposed to *M. tuberculosis* receive chest x-rays and attempts are made to culture the bacterium *in vitro* from sputum samples. X-rays are insensitive as a diagnostic assay and can only identify infections in a very progressed stage. Culturing of *M. tuberculosis* is also not ideal as a diagnostic tool, since the bacteria grows poorly and slowly outside the body, which can produce false negative test results and take weeks before results are obtained. The standard tuberculin skin test is an inexpensive assay, used in third world countries, however it is far from ideal in detecting infection because it cannot distinguish *M. tuberculosis*-infected individuals from *M. bovis* BCG-vaccinated individuals and therefore cannot be used in areas of the world where patients receive or have received childhood vaccination with bacterial strains related to *M. tuberculosis*, e.g. a BCG vaccination.

Animal tuberculosis is caused by *Mycobacterium bovis*, which is closely related to *M. tuberculosis* and within the tuberculosis complex. *M. bovis* is an important pathogen that can infect a range of hosts, including cattle and humans. Tuberculosis in cattle is a major cause of economic loss and represents a significant cause of zoonotic infection. A number of strategies have been employed against bovine TB, but the approach has generally been based on government-organised programmes by which animals deemed positive to defined screening test are slaughtered. The most common test used in cattle is Delayed-type hypersensitivity with PPD as antigen, but alternative *in vitro* assays are also developed. However, investigations have shown the both the *in vivo* and the *in vitro* tests have a relative low specificity, and the detection of false-positive is a significant economic problem (Pollock et al 2000). There is therefore a great need for a more specific diagnostic reagent, which can be used either *in vivo* or *in vitro* to detect *M. bovis* infections in animals.

Summary of the invention

The present invention is related to preventing, treating and detecting infections caused by species of the tuberculosis complex (*M. tuberculosis*, *M.bovis*, *M. africanum*) by the use of a polypeptide comprising an immunogenic portion of one or more of the polypeptides TB10.3 (also named ORF7-1 or Rv3019c), TB10.4 (also named CFP7 or Rv0288) and TB12.9 (also named ORF7-2 or Rv3017c) (WO98/44119, WO99/24577 and Skjøt et al, 2000) or by a nucleotide sequence comprising a nucleotide sequence encoding an immunogenic portion of TB10.3, TB10.4 or TB12.9.

Detailed disclosure

The present invention discloses a substantially pure polypeptide, which comprises an amino acid sequence selected from

(a) the group consisting of Rv0288 (SEQ ID NO: 2 and 195) and its homologues Rv3019c (SEQ ID NO: 199) and Rv3017c (SEQ ID NO: 197);

(b) an immunogenic portion, e.g. a T-cell epitope, of any one of the sequences in (a); and /or

(c) an amino acid sequence analogue having at least 70% sequence identity to any one of the sequences in (a) or (b) and at the same time being immunogenic.

Preferred immunogenic portions are the fragments TB10.3-P1, TB10.3-P2, TB10.3-P3, TB10.3-P4, TB10.3-P5, TB10.3-P6, TB10.3-P7, TB10.3-P8, TB10.3-P9, TB10.4-P1, TB10.4-P2, TB10.4-P3, TB10.4-P4, TB10.4-P5, TB10.4-P6, TB10.4-P7, TB10.4-P8, TB10.4-P9, TB12.9-P1, TB12.9-P2, TB12.9-P3, TB12.9-P4, TB12.9-P5, TB12.9-P6, TB12.9-P7, TB12.9-P8, TB12.9-P9, TB12.9-P10 and TB12.9-P11, which have immunological activity. They are recognized in an *in vitro* cellular assay determining the release of IFN-γ from lymphocytes withdrawn from an individual currently or previously infected with a virulent mycobacterium.

Further, the present invention discloses a vaccine, a pharmaceutical composition and a diagnostic reagent, all comprising an amino acid sequence selected from

(a) the group consisting of Rv0288 (SEQ ID NO: 2 and 195) and its homologues Rv3019c (SEQ ID NO: 199) or Rv3017c (SEQ ID NO: 197);

(b) an immunogenic portion, e.g. a T-cell epitope, of any one of the sequences in (a); and /or

(c) an amino acid sequence analogue having at least 70% sequence identity to any one of the sequences in (a) or (b) and at the same time being immunogenic.

Definitions

The word "polypeptide" in the present specification and claims should have its usual meaning. That is an amino acid chain of any length, including a full-length protein, oligopeptides, short peptides and fragments thereof, wherein the amino acid residues are linked by covalent peptide bonds.

The polypeptide may be chemically modified by being glycosylated, by being lipidated (e.g. by chemical lipidation with palmitoyloxy succinimide as described by Mowat et al. 1991 or with dodecanoyl chloride as described by Lustig et al. 1976), by comprising prosthetic groups, or by containing additional amino acids such as e.g. a his-tag or a signal peptide.

Each polypeptide may thus be characterised by comprising specific amino acid sequences and be encoded by specific nucleic acid sequences. It will be understood that such sequences include analogues and variants produced by recombinant or synthetic methods wherein such polypeptide sequences have been modified by substitution, insertion, addition or deletion of one or more amino acid residues in the recombinant polypeptide and still be immunogenic in any of the biological assays described herein. Substitutions are preferably "conservative". These are defined according to the following table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other. The amino acids in the third column are indicated in one-letter code.

ALIPHATIC	Non-polar	GAP
		ILV
	Polar-uncharged	CSTM
		NQ
	Polar-charged	DE
		KR
AROMATIC		HFWY

A preferred polypeptide within the present invention is an immunogenic antigen from *M. tuberculosis*. Such antigen can for example be derived from *M. tuberculosis* and/or *M. tuberculosis* culture filtrate. Thus, a polypeptide comprising an immunogenic portion of one of the above antigens may consist entirely of the immunogenic portion, or may contain additional sequences. The additional sequences may be derived from the native *M. tuberculosis* antigen or be heterologous and such sequences may, but need not, be immunogenic.

Each polypeptide is encoded by a specific nucleic acid sequence. It will be understood that such sequences include analogues and variants hereof wherein such nucleic acid sequences have been modified by substitution, insertion, addition or deletion of one or more nucleic acids. Substitutions are preferably silent substitutions in the codon usage which will not lead to any change in the amino acid sequence, but may be introduced to enhance the expression of the protein.

In the present context the term "substantially pure polypeptide fragment" means a polypeptide preparation which contains at most 5% by weight of other polypeptide material with which it is natively associated (lower percentages of other polypeptide material are preferred, e.g. at most 4%, at most 3%, at most 2%, at most 1%, and at most ½%). It is preferred that the substantially pure polypeptide is at least 96% pure, *i.e.* that the polypeptide constitutes at least 96% by weight of total polypeptide material present in the preparation, and higher percentages are preferred, such as at least 97%, at least 98%, at least 99%, at least 99,25%, at least 99,5%, and at least 99,75%. It is especially preferred that the polypeptide fragment is in "essentially pure form", *i.e.* that the polypeptide fragment is essentially free of any other antigen with which it is natively associated, *i.e.* free of any other antigen from bacteria belonging to the tuberculosis complex or a virulent mycobacterium. This can be accomplished by preparing the polypeptide fragment by means of recombinant methods in a non-mycobacterial host cell as will be described in detail below, or by synthesizing the polypeptide fragment by the well-known methods of solid or liquid phase peptide synthesis, e.g. by the method described by Merrifield or variations thereof.

By the term "virulent mycobacterium" is understood a bacterium capable of causing the tuberculosis disease in an animal or in a human being. Examples of virulent mycobacteria are *M. tuberculosis, M. africanum,* and *M. bovis.* Examples of relevant animals are cattle, possums, badgers and kangaroos.

By "a TB patient" is understood an individual with culture or microscopically proven infection with virulent mycobacteria, and/or an individual clinically diagnosed with TB and who is responsive to anti-TB chemotherapy. Culture, microscopy and clinical diagnosis of TB are well known by any person skilled in the art.

By the term "PPD-positive individual" is understood an individual with a positive Mantoux test or an individual where PPD induces a positive *in vitro* recall response determined by release of IFN-γ.

By the term "delayed type hypersensitivity reaction" (DTH) is understood a T-cell mediated inflammatory response elicited after the injection of a polypeptide into, or application to, the skin, said inflammatory response appearing 72-96 hours after the polypeptide injection or application.

By the term "IFN-γ" is understood interferon-gamma. The measurement of IFN-γ is used as an indication of an immunological response.

By the terms "nucleic acid fragment" and "nucleic acid sequence" are understood any nucleic acid molecule including DNA, RNA , LNA (locked nucleic acids), PNA, RNA, dsRNA and RNA-DNA-hybrids. Also included are nucleic acid molecules comprising non-naturally occurring nucleosides. The term includes nucleic acid molecules of any length e.g. from 10 to 10000 nucleotides, depending on the use. When the nucleic acid molecule is for use as a pharmaceutical, e.g. in DNA therapy, or for use in a method for producing a polypeptide according to the invention, a molecule encoding at least one epitope is preferably used, having a length from about 18 to about 1000 nucleotides, the molecule being optionally inserted into a vector. When the nucleic acid molecule is used as a probe, as a primer or in antisense therapy, a molecule having a length of 10-100 is preferably used. According to the invention, other molecule lengths can be used, for instance a molecule having at least 12, 15, 21, 24, 27, 30, 33, 36, 39, 42, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 or 1000 nucleotides (or nucleotide

derivatives), or a molecule having at most 10000, 5000, 4000, 3000, 2000, 1000, 700, 500, 400, 300, 200, 100, 50, 40, 30 or 20 nucleotides (or nucleotide derivatives).

The term "stringent" when used in conjunction with nucleic acid hybridization conditions is as defined in the art, i.e. the hybridization is performed at a temperature not more than 15-20°C under the melting point Tm, cf. Sambrook et al, 1989, pages 11.45-11.49. Preferably, the conditions are "highly stringent", i.e. 5-10°C under the melting point Tm.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations thereof such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

The term "sequence identity" indicates a quantitative measure of the degree of homology between two amino acid sequences of equal length or between two nucleotide sequences of equal length. If the two sequences to be compared are not of equal length, they must be aligned to best possible fit possible with the insertion of gaps or alternatively, truncation at the ends of the protein sequences. The sequence identity can be calculated as $\frac{(N_{ref}-N_{dif})100}{N_{ref}}$, wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of residues in one of the sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC (N_{dif}=2 and N_{ref}=8). A gap is counted as non-identity of the specific residue(s), i.e. the DNA sequence AGTGTC will have a sequence identity of 75% with the DNA sequence AGTCAGTC (N_{dif}=2 and N_{ref}=8). Sequence identity can alternatively be calculated by the BLAST program e.g. the BLASTP program (Pearson W.R and D.J. Lipman (1988) PNAS USA 85:2444-2448)(www.ncbi.nlm.nih.gov/cgi-bin/BLAST). In one aspect of the invention, alignment is performed with the sequence alignment method ClustalW with default parameters as described by Thompson J., *et al* 1994, available at http://www2.ebi.ac.uk/clustalw/.

A preferred minimum percentage of sequence identity is at least 70%, such as at least 75%, at least 80%, such as at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, such as at least 99.5%.

In a preferred embodiment of the invention, the polypeptide comprises an immunogenic portion of the polypeptide, such as an epitope for a B-cell or T-cell. The immunogenic portion of a polypeptide is a part of the polypeptide, which elicits an immune response in an animal or a human being, and/or in a biological sample determined by any of the biological assays described herein. The immunogenic portion of a polypeptide may be a T-cell epitope or a B-cell epitope. Immunogenic portions can be related to one or a few relatively small parts of the polypeptide, they can be scattered throughout the polypeptide sequence or be situated in specific parts of the polypeptide. For a few polypeptides epitopes have even been demonstrated to be scattered throughout the polypeptide covering the full sequence (Ravn et al 1999).

In order to identify relevant T-cell epitopes which are recognised during an immune response, it is possible to use a "brute force" method: Since T-cell epitopes are linear, deletion mutants of the polypeptide will, if constructed systematically, reveal what regions of the polypeptide are essential in immune recognition, e.g. by subjecting these deletion mutants e.g. to the IFN-γ assay described herein. Another method utilises overlapping oligopeptides for the detection of MHC class II epitopes, preferably synthetic, having a length of e.g. 20 amino acid residues derived from the polypeptide. These peptides can be tested in biological assays (e.g. the IFN-γ assay as described herein) and some of these will give a positive response (and thereby be immunogenic) as evidence for the presence of a T cell epitope in the peptide. For the detection of MHC class I epitopes it is possible to predict peptides that will bind (Stryhn et al 1996) and hereafter produce these peptides synthetically and test them in relevant biological assays e.g. the IFN-γ assay as described herein. The peptides preferably having a length of e.g. 8 to 11 amino acid residues derived from the polypeptide. B-cell epitopes can be determined by analysing the B cell recognition to overlapping peptides covering the polypeptide of interest as e.g. described in Harboe et al 1998.

Although the minimum length of a T-cell epitope has been shown to be at least 6 amino acids, it is normal that such epitopes are constituted of longer stretches of amino acids. Hence, it is preferred that the polypeptide fragment of the invention has a length of at least 7 amino acid residues, such as at least 8, at least 9, at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 22, at least 24, such as at least 30 amino acid residues. Hence, in important embodiments of the inventive method, it is preferred that the polypeptide fragment has a length of at most 50 amino acid residues, such as at most 40, 35, 30, 25, e.g. at most 20

amino acid residues. It is expected that the peptides having a length of between 10 and 20 amino acid residues will prove to be most efficient as MHC class II epitopes and therefore especially preferred lengths of the polypeptide fragment used in the method according to the invention are 18, such as 15, 14, 13, 12 and even 11 amino acid residues. It is expected that the peptides having a length of between 7 and 12 amino acid residues will prove to be most efficient as MHC class I epitopes and therefore especially other lengths of the polypeptide fragment used in the method according to the invention are 11, such as 10, 9, 8 and even 7 amino acid residues.

Immunogenic portions of polypeptides may be recognised by a broad part (high frequency) or by a minor part (low frequency) of the genetically heterogenic human population. In addition some immunogenic portions induce high immunological responses (dominant), whereas others induce lower, but still significant, responses (subdominant). High frequency><low frequency can be related to the immunogenic portion binding to widely distributed MHC molecules (HLA type) or even by multiple MHC molecules (Kilgus et al. 1991, Sinigaglia et al 1988).

In the context of providing candidate molecules for a new vaccine against tuberculosis, the subdominant epitopes are however as relevant as are the dominant epitopes since it has been show (Olsen et al 2000) that such epitopes can induce protection regardless of being subdominant.

A common feature of the polypeptides of the invention is their capability to induce an immunological response as illustrated in the examples. It is understood that a variant of a polypeptide of the invention produced by substitution, insertion, addition or deletion is also immunogenic as determined by at least one of the assays described herein.

An immune individual is defined as a person or an animal, which has cleared or controlled an infection with virulent mycobacteria or has received a vaccination with *M.bovis* BCG.

An immunogenic polypeptide is defined as a polypeptide that induces an immune response in a biological sample or an individual currently or previously infected with a virulent mycobacterium. The immune response may be monitored by one of the following methods:

- An *in vitro* cellular response is determined by release of a relevant cytokine such as IFN-γ from lymphocytes withdrawn from an animal or human being currently or previously infected with virulent mycobacteria, or by detection of proliferation of these T cells, the induction being performed by the addition of the polypeptide or the immunogenic portion to a suspension comprising from $1x10^5$ cells to $3x10^5$ cells per well. The cells are isolated from either the blood, the spleen, the liver or the lung and the addition of the polypeptide or the immunogenic portion resulting in a concentration of not more than 20 μg per ml suspension and the stimulation being performed from two to five days. For monitoring cell proliferation the cells are pulsed with radioactive labeled Thymidine and after 16-22 hours of incubation detecting the proliferation by liquid scintillation counting, a positive response being a response more than background plus two standard derivations. The release of IFN-γ can be determined by the ELISA method, which is well known to a person skilled in the art, a positive response being a response more than background plus two standard derivations. Other cytokines than IFN-γ could be relevant when monitoring the immunological response to the polypeptide, such as IL-12, TNF-α, IL-4, IL-5, IL-10, IL-6, TGF-β. Another and more sensitive method for determining the presence of a cytokine (e.g. IFN-γ) is the ELISPOT method where the cells isolated from either the blood, the spleen, the liver or the lung are diluted to a concentration of preferably 1 to 4 x 10^6 cells /ml and incubated for 18-22 hrs in the presence of the polypeptide or the immunogenic portion resulting in a concentration of not more than 20 μg per ml. The cell suspensions are hereafter diluted to 1 to 2 x 10^6/ ml and transferred to Maxisorp plates coated with anti–IFN-γ and incubated for preferably 4 to 16 hours. The IFN-γ producing cells are determined by the use of labelled secondary anti-IFN-γ antibody and a relevant substrate giving rise to spots, which can be enumerated using a dissection microscope. It is also a possibility to determine the presence of mRNA coding for the relevant cytokine by the use of the PCR technique. Usually one or more cytokines will be measured utilizing for example the PCR, ELISPOT or ELISA. It will be appreciated by a person skilled in the art that a significant increase or decrease in the amount of any of these cytokines induced by a specific polypeptide can be used in evaluation of the immunological activity of the polypeptide.

- An *in vitro* cellular response may also be determined by the use of T cell lines derived from an immune individual or a person infected with *M. tuberculosis* where the T cell

lines have been driven with either live mycobacteria, extracts from the bacterial cell or culture filtrate for 10 to 20 days with the addition of IL-2. The induction is performed by addition of not more than 20 µg polypeptide per ml suspension to the T cell lines containing from $1x10^5$ cells to $3x10^5$ cells per well and incubation being performed from two to six days. The induction of IFN-γ or release of another relevant cytokine is detected by ELISA. The stimulation of T cells can also be monitored by detecting cell proliferation using radioactively labeled Thymidine as described above. For both assays a positive response is a response more than background plus two standard derivations.

- An *in vivo* cellular response which may be determined as a positive DTH response after intradermal injection or local application patch of at most 100 µg of the polypeptide or the immunogenic portion to an individual who is clinically or subclinically infected with a virulent Mycobacterium, a positive response having a diameter of at least 5 mm 72-96 hours after the injection or application.

- An *in vitro* humoral response is determined by a specific antibody response in an immune or infected individual. The presence of antibodies may be determined by an ELISA technique or a Western blot where the polypeptide or the immunogenic portion is absorbed to either a nitrocellulose membrane or a polystyrene surface. The serum is preferably diluted in PBS from 1:10 to 1:100 and added to the absorbed polypeptide and the incubation being performed from 1 to 12 hours. By the use of labeled secondary antibodies the presence of specific antibodies can be determined by measuring the OD e.g. by ELISA where a positive response is a response of more than background plus two standard derivations or alternatively a visual response in a Western blot.

- Another relevant parameter is measurement of the protection in animal models induced after vaccination with the polypeptide in an adjuvant or after DNA vaccination. Suitable animal models include primates, guinea pigs or mice, which are challenged with an infection of a virulent Mycobacterium. Readout for induced protection could be decrease of the bacterial load in target organs compared to non-vaccinated animals, prolonged survival times compared to non-vaccinated animals and diminished weight loss compared to non-vaccinated animals.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any one of a variety of procedures. They may be purified as native proteins from the *M. tuberculosis* cell or culture filtrate by procedures such as those described above. Immunogenic antigens may also be produced recombinantly using a DNA sequence encoding the antigen, which has been inserted into an expression vector and expressed in an appropriate host. Examples of host cells are *E. coli*. The polypeptides or immunogenic portion hereof can also be produced synthetically if having fewer than about 100 amino acids, generally fewer than 50 amino acids, and may be generated using techniques well known to those ordinarily skilled in the art, such as commercially available solid-phase techniques where amino acids are sequentially added to a growing amino acid chain.

In the construction and preparation of plasmid DNA encoding the polypeptide as defined for DNA vaccination a host strain such as *E. coli* can be used. Plasmid DNA can then be prepared from overnight cultures of the host strain carrying the plasmid of interest and purified using e.g. the Qiagen Giga -Plasmid column kit (Qiagen, Santa Clarita, CA, USA) including an endotoxin removal step. It is essential that plasmid DNA used for DNA vaccination is endotoxin free.

The immunogenic polypeptides may also be produced as fusion proteins, by which methods superior characteristics of the polypeptide of the invention can be achieved. For instance, fusion partners that facilitate export of the polypeptide when produced recombinantly, fusion partners that facilitate purification of the polypeptide, and fusion partners which enhance the immunogenicity of the polypeptide fragment of the invention are all interesting possibilities. Therefore, the invention also pertains to a fusion polypeptide comprising at least one polypeptide or immunogenic portion defined above and at least one fusion partner. The fusion partner can, in order to enhance immunogenicity, be another polypeptide derived from *M. tuberculosis*, such as of a polypeptide fragment derived from a bacterium belonging to the tuberculosis complex, such as ESAT-6, TB10.4, CFP10, RD1-ORF5, RD1-ORF2, Rv1036, MPB64, MPT64, Ag85A, Ag85B (MPT59), MPB59, , Ag85C, 19kDa lipoprotein, MPT32 and alpha-crystallin, or at least one T-cell epitope of any of the above mentioned antigens ((Skjøt et al, 2000; Danish Patent application PA 2000 00666; Danish Patent application PA 1999 01020; US patent application 09/0505,739; Rosenkrands et al, 1998; Nagai et al, 1991). The invention also pertains to a fusion polypeptide comprising mutual fusions of two or more of the polypeptides (or immunogenic portions thereof) of the invention.

Other fusion partners, which could enhance the immunogenicity of the product, are lymphokines such as IFN-γ, IL-2 and IL-12. In order to facilitate expression and/or purification, the fusion partner can e.g. be a bacterial fimbrial protein, e.g. the pilus components pilin and papA; protein A; the ZZ-peptide (ZZ-fusions are marketed by Pharmacia in Sweden); the maltose binding protein; gluthatione S-transferase; β-galactosidase; or poly-histidine. Fusion proteins can be produced recombinantly in a host cell, which could be *E. coli*, and it is a possibility to induce a linker region between the different fusion partners.

Other interesting fusion partners are polypeptides, which are lipidated so that the immunogenic polypeptide is presented in a suitable manner to the immune system. This effect is e.g. known from vaccines based on the Borrelia burgdorferi OspA polypeptide as described in e.g. WO 96/40718 A or vaccines based on the Pseudomonas aeruginosa OprI lipoprotein (Cote-Sierra J, et al, 1998). Another possibility is N-terminal fusion of a known signal sequence and an N-terminal cystein to the immunogenic polypeptide. Such a fusion results in lipidation of the immunogenic polypeptide at the N-terminal cystein, when produced in a suitable production host.

Another part of the invention pertains to a vaccine composition comprising a polypeptide (or at least one immunogenic portion thereof) or fusion polypeptide according to the invention. In order to ensure optimum performance of such a vaccine composition it is preferred that it comprises an immunologically and pharmaceutically acceptable carrier, vehicle or adjuvant.

An effective vaccine, wherein a polypeptide of the invention is recognized by the animal, will in an animal model be able to decrease bacterial load in target organs, prolong survival times and/or diminish weight loss after challenge with a virulent Mycobacterium, compared to non-vaccinated animals.

Suitable carriers are selected from the group consisting of a polymer to which the polypeptide(s) is/are bound by hydrophobic non-covalent interaction, such as a plastic, e.g. polystyrene, or a polymer to which the polypeptide(s) is/are covalently bound, such as a polysaccharide, or a polypeptide, e.g. bovine serum albumin, ovalbumin or keyhole limpet haemocyanin. Suitable vehicles are selected from the group consisting of a diluent and a suspending agent. The adjuvant is preferably selected from the group consisting of dimethyldioctadecylammonium

bromide (DDA), Quil A, poly I:C, aluminium hydroxide, Freund's incomplete adjuvant, IFN-γ, IL-2, IL-12, monophosphoryl lipid A (MPL), Treholose Dimycolate (TDM), Trehalose Dibehenate and muramyl dipeptide (MDP).

Preparation of vaccines which contain peptide sequences as active ingredients is generally well understood in the art, as exemplified by U.S. Patents 4,608,251; 4,601,903; 4,599,231 and 4,599,230, all incorporated herein by reference.

Other methods of achieving adjuvant effect for the vaccine include use of agents such as aluminum hydroxide or phosphate (alum), synthetic polymers of sugars (Carbopol), aggregation of the protein in the vaccine by heat treatment, aggregation by reactivating with pepsin treated (Fab) antibodies to albumin, mixture with bacterial cells such as C. parvum or endotoxins or lipopolysaccharide components of gram-negative bacteria, emulsion in physiologically acceptable oil vehicles such as mannide mono-oleate (Aracel A) or emulsion with 20 percent solution of a perfluorocarbon (Fluosol-DA) used as a block substitute may also be employed. Other possibilities involve the use of immune modulating substances such as cytokines or synthetic IFN-γ inducers such as poly I:C in combination with the above-mentioned adjuvants.

Another interesting possibility for achieving adjuvant effect is to employ the technique described in Gosselin *et al.*, 1992 (which is hereby incorporated by reference herein). In brief, a relevant antigen such as an antigen of the present invention can be conjugated to an antibody (or antigen binding antibody fragment) against the Fcγ receptors on monocytes/macrophages.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective and immunogenic. The quantity to be administered depends on the subject to be treated, including, e.g., the capacity of the individual's immune system to mount an immune response, and the degree of protection desired. Suitable dosage ranges are of the order of several hundred micrograms active ingredient per vaccination with a preferred range from about 0.1 μg to 1000 μg, such as in the range from about 1 μg to 300 μg, and especially in the range from about 10 μg to 50 μg. Suitable regimens for initial administration and booster shots are also variable but are typified by an initial administration followed by subsequent inoculations or other administrations.

The manner of application may be varied widely. Any of the conventional methods for administration of a vaccine are applicable. These are believed to include oral application on a solid physiologically acceptable base or in a physiologically acceptable dispersion, parenterally, by injection or the like. The dosage of the vaccine will depend on the route of administration and will vary according to the age of the person to be vaccinated and, to a lesser degree, the size of the person to be vaccinated.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkalene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and advantageously contain 10-95% of active ingredient, preferably 25-70%.

In many instances, it will be necessary to have multiple administrations of the vaccine. Especially, vaccines can be administered to prevent an infection with virulent mycobacteria and/or to treat established mycobacterial infection. When administered to prevent an infection, the vaccine is given prophylactically, before definitive clinical signs or symptoms of an infection are present.

Due to genetic variation, different individuals may react with immune responses of varying strength to the same polypeptide. Therefore, the vaccine according to the invention may comprise several different polypeptides in order to increase the immune response. The vaccine may comprise two or more polypeptides or immunogenic portions, where all of the polypeptides are as defined above, or some but not all of the peptides may be derived from virulent mycobacteria. In the latter example, the polypeptides not necessarily fulfilling the criteria set forth above for polypeptides may either act due to their own immunogenicity or merely act as adjuvants.

The vaccine may comprise 1-20, such as 2-20 or even 3-20 different polypeptides or fusion polypeptides, such as 3-10 different polypeptides or fusion polypeptides.

The invention also pertains to a method for immunising an animal, including a human being, against TB caused by virulent mycobacteria, comprising administering to the animal the polypeptide of the invention, or a vaccine composition of the invention as described above, or a living vaccine described above.

The invention also pertains to a method for producing an immunologic composition according to the invention, the method comprising preparing, synthesising or isolating a polypeptide according to the invention, and solubilizing or dispersing the polypeptide in a medium for a vaccine, and optionally adding other *M. tuberculosis* antigens and/or a carrier, vehicle and/or adjuvant substance.

The nucleic acid fragments of the invention may be used for effecting *in vivo* expression of antigens, *i.e.* the nucleic acid fragments may be used in so-called DNA vaccines as reviewed in Ulmer et al., 1993, which is included by reference.

Hence, the invention also relates to a vaccine comprising a nucleic acid fragment according to the invention, the vaccine effecting *in vivo* expression of antigen by an animal, including a human being, to whom the vaccine has been administered, the amount of expressed antigen being effective to confer substantially increased resistance to infections caused by virulent mycobacteria in an animal, including a human being.

The efficacy of such a DNA vaccine can possibly be enhanced by administering the gene encoding the expression product together with a DNA fragment encoding a polypeptide which has the capability of modulating an immune response.

One possibility for effectively activating a cellular immune response for a vaccine can be achieved by expressing the relevant antigen in a vaccine in a non-pathogenic microorganism or virus. Well-known examples of such microorganisms are *Mycobacterium bovis* BCG, *Salmonella* and *Pseudomona* and examples of viruses are Vaccinia Virus and Adenovirus.

Therefore, another important aspect of the present invention is an improvement of the living BCG vaccine presently available, wherein one or more copies of a DNA sequence encoding one or more polypeptide as defined above has been incorporated into the genome of the micro-organism in a manner allowing the micro-organism to express and secrete the polypeptide. The incorporation of more than one copy of a nucleotide sequence of the invention is contemplated to enhance the immune response

Another possibility is to integrate the DNA encoding the polypeptide according to the invention in an attenuated virus such as the vaccinia virus or Adenovirus (Rolph et al 1997). The recombinant vaccinia virus is able to replicate within the cytoplasma of the infected host cell and the polypeptide of interest can therefore induce an immune response, which is envisioned to induce protection against TB.

The invention also relates to the use of a polypeptide or nucleic acid of the invention for use as a therapeutic vaccine, which concept has been described in the literature exemplified by D. Lowry (1999, Nature 400: 269-71). Antigens with therapeutic properties may be identified based on their ability to diminish the severity of *M. tuberculosis* infection in experimental animals or prevent reactivation of previous infection, when administered as a vaccine. The composition used for therapeutic vaccines can be prepared as described above for vaccines.

The invention also relates to a method of diagnosing TB caused by a virulent mycobacterium in an animal, including a human being, comprising intradermally injecting, in the animal, a polypeptide according to the invention, a positive skin response at the location of injection being indicative of the animal having TB, and a negative skin response at the location of injection being indicative of the animal not having TB.

When diagnosis of previous or ongoing infection with virulent mycobacteria is the aim, a blood sample comprising mononuclear cells (*i.e.* T-lymphocytes) from a patient could be contacted with a sample of one or more polypeptides of the invention. This contacting can be performed *in vitro* and a positive reaction could e.g. be proliferation of the T-cells or release of cytokines such as IFN-γ into the extracellular phase. It is also conceivable to contact a serum sample from a subject with a polypeptide of the invention, the demonstration of a binding between antibodies in the serum sample and the polypeptide being indicative of previous or ongoing infection.

The invention therefore also relates to an *in vitro* method for diagnosing ongoing or previous sensitisation in an animal or a human being with a virulent mycobacterium, the method comprising providing a blood sample from the animal or human being, and contacting the sample from the animal with the polypeptide of the invention, a significant release into the extracellular phase of at least one cytokine by mononuclear cells in the blood sample being indicative of the animal being sensitised. A positive response is a response more than release from a blood sample derived from a patient without the TB diagnosis plus two standard derivations. The invention also relates to an *in vitro* method for diagnosing ongoing or previous sensitisation in an animal or a human being with a virulent mycobacterium, the method comprising providing a blood sample from the animal or human being, and contacting the sample from the animal with the polypeptide of the invention demonstrating the presence of antibodies recognizing the polypeptide of the invention in the serum sample.

The immunogenic composition used for diagnosing may comprise 1-20, such as 2-20 or even 3-20 different polypeptides or fusion polypeptides, such as 3-10 different polypeptides or fusion polypeptides.

The nucleic acid probes encoding the polypeptide of the invention can be used in a variety of diagnostic assays for detecting the presence of pathogenic organisms in a given sample. A method of determining the presence of mycobacterial nucleic acids in an animal, including a human being, or in a sample, comprising administering a nucleic acid fragment of the invention to the animal or incubating the sample with the nucleic acid fragment of the invention or a nucleic acid fragment complementary thereto, and detecting the presence of hybridised nucleic acids resulting from the incubation (by using the hybridisation assays which are well-known in the art), is also included in the invention. Such a method of diagnosing TB might involve the use of a composition comprising at least a part of a nucleotide sequence as defined above and detecting the presence of nucleotide sequences in a sample from the animal or human being to be tested which hybridise with the nucleic acid fragment (or a complementary fragment) by the use of PCR technique.

A monoclonal or polyclonal antibody, which is specifically reacting with a polypeptide of the invention in an immuno assay, or a specific binding fragment of said antibody, is also a part of the invention. The antibodies can be produced by methods known to the person skilled in the art. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of

a polypeptide according to the present invention and, if desired, an adjuvant. The monoclonal antibodies according to the present invention may, for example, be produced by the hybridoma method first described by Köhler and Milstein (1975), or may be produced by recombinant DNA methods such as described in US patent no. 4,816,567. The monoclonal antibodies may also be isolated from phage libraries generated using the techniques described in McCafferty et al (1990), for example. Methods for producing antibodies are described in the literature, e.g. in US patent no. 6,136,958.

A sample of a potentially infected organ may be contacted with such an antibody recognizing a polypeptide of the invention. The demonstration of the reaction by means of methods well known in the art between the sample and the antibody will be indicative of an ongoing infection. It is of course also a possibility to demonstrate the presence of anti-mycobacterial antibodies in serum by contacting a serum sample from a subject with at least one of the polypeptide fragments of the invention and using well-known methods for visualising the reaction between the antibody and antigen.

In diagnostics, an antibody, a nucleic acid fragment and/or a polypeptide of the invention can be used either alone, or as a constituent in a composition. Such compositions are known in the art, and comprise compositions in which the antibody, the nucleic acid fragment or the polypeptide of the invention is coupled, preferably covalently, to at least one other molecule, e.g. a label (e.g. radioactive or fluorescent) or a carrier molecule.

Concordance list

	Synonyms	DNA SEQ ID NO	Protein SEQ ID NO
Rv0288	CFP7,TB10.4, PV-2 binding protein	1 and 194	2 and 195
Rv3017c	ORF7-2,TB12.9	196	197
Rv3019c	ORF7-1,TB10.3	198	199
TB10.3-P1		200	201
TB10.3-P2		202	203
TB10.3-P3		204	205
TB10.3-P4		206	207
TB10.3-P5		208	209
TB10.3-P6		210	211
TB10.3-P7		212	213
TB10.3-P8		214	215
TB10.3-P9		216	217
TB10.4-P1		218	219
TB10.4-P2		220	221
TB10.4-P3		222	223
TB10.4-P4		224	225
TB10.4-P5		226	227
TB10.4-P6		228	229
TB10.4-P7		230	231
TB10.4-P8		232	233
TB10.4-P9		234	235
TB12.9-P1		236	237
TB12.9-P2		238	239
TB12.9-P3		240	241
TB12.9-P4		242	243
TB12.9-P5		244	245
TB12.9-P6		246	247
TB12.9-P7		248	249
TB12.9-P8		250	251
TB12.9-P9		252	253
TB12.9-P10		254	255
TB12.9-P11		256	257

Legends to figures

Fig. 1: Course of infection with *M. tuberculosis* in naive and memory immune mice.
C57Bl/6j mice were infected with 2.5 x 10^5 viable units of *M. tuberculosis* and the growth of the organisms in the spleen was investigated for a period of 25 days. The count of the CFU indicated represent the means of 4-5 mice.

Fig. 2: *In vivo* IFN-γ production during tuberculosis infection.
Memory immune or naive mice were infected with 2.5 x 10^5 colony forming units of *M. tuberculosis* i.v. and the level of IFN-γ was monitored in the spleen or serum of animals during the course of infection.

Fig. 3: *In vitro* response of spleen lymphocytes from infected mice.
Memory immune or naive mice were sacrificed at different time points during the course of infection, and spleen lymphocytes were stimulated *in vitro* with ST-CF or killed bacilli. Cell culture supernatants were tested for the presence of IFN-γ.

Fig. 4: Short-term culture-filtrate fractions.
ST-CF was divided into 14 fractions by the multi-elution technique and the fractions were analyzed by SDS-PAGE and Silver-staining. Lane F: ST-CF Lane 1-15: fractions 1-15.

Fig. 5: T-cell reactivation during a secondary infection.
IFN-γ release by spleen lymphocytes isolated either directly from memory immune mice or four days after the mice had received a secondary infection. The lymphocytes were stimulated *in vitro* with ST-CF fractions and the supernatants harvested for quantification of IFN-γ. The migration of molecular mass markers (as shown in Fig. 4) are indicated at the bottom.

Fig. 6: Precise mapping of IFN-γ release in response to single secreted antigens.
A panel of narrow fractions within the stimulatory regions 4-14 and 26-34 enabled the precise mapping of proteins capable of inducing IFN-γ in microcultures containing lymphocytes from memory immune mice at day 4 of rechallenge.
On the left hand side: IFN-γ release by single secreted antigens.

On the right hand side: The localization of and IFN-γ induction by defined secreted antigens of *M. tuberculosis*. ST-3, 76-8 and PV-2 are the designation of three mAbs which defines secreted antigens of molecular mass 5-8 kDa.

Fig. 7: Physical map of recombinant lambda phages expressing products reactive with Mabs recognizing low MW components.
Cross-hatched bar; *lacZ*, solid bar; *M. tuberculosis* DNA, open bar; lambdagt11 DNA (right arm), open triangles indicate *EcoR*I cleavage sites originating from the lambdagt11 vector. The direction of translation and transcription of the gene products fused to beta-galactosidase is indicated by an arrow.

Fig. 8: Western blot analyses demonstrating recombinant expression of low molecular weight components.
Lysates of *E. coli* Y1089 lysogenized with lambda AA226, lambda AA227 or lambda were analyzed in Western blot experiments after PAGE (A: 10%, B: 10 to 20% gradient).
Panel A: lanes 1: lambda gt11, lanes 2: lambda AA226, lanes 3: lambda AA227.
Panel B: lane 1: lambda gt11, lanes 2 and 3: lambda AA242 and AA230 (identical clones).
The monoclonal antibodies are indicated on top of each panel. L24,c24 is an anti-MPT64 reactive monoclonal antibody.

Fig. 9: Nucleotide sequence (SEQ ID NO: 1) of *cfp7*. The deduced amino acid sequence (SEQ ID NO: 2) of CFP7 is given in conventional one-letter code below the nucleotide sequence. The putative ribosome-binding site is written in underlined italics as are the putative -10 and -35 regions. Nucleotides written in bold are those encoding CFP7.

Fig. 10. Nucleotide sequence (SEQ ID NO: 3) of *cfp9*. The deduced amino acid sequence (SEQ ID NO: 4) of CFP9 is given in conventional one-letter code below the nucleotide sequence. The putative ribosome-binding site Shine Delgarno sequence is written in underlined italics as are the putative -10 and -35 regions. Nucleotides in bold writing are those encoding CFP9. The nucleotide sequence obtained from the lambda 226 phage is double underlined.

Fig. 11: Nucleotide sequence of *mpt51*. The deduced amino acid sequence of MPT51 is given in a one-letter code below the nucleotide sequence. The signal is indicated in italics. The

putative potential ribosome-binding site is underlined. The nucleotide difference and amino acid difference compared to the nucleotide sequence of MPB51 (Ohara *et al.*, 1995) are underlined at position 780. The nucleotides given in italics are not present in *M. tuberculosis* H37Rv.

Fig. 12: the position of the purified antigens in the 2DE system have been determined and mapped in a reference gel. The newly purified antigens are encircled and the position of well-known proteins are also indicated.

Figure 13
Indication of the TB10.4 immunogenic portions in alignment to the full sequence of TB10.4.

Figure 14
Indication of the TB10.3 immunogenic portions in alignment to the full sequence of TB10.3. Underlined amino acids are different from the TB10.4 peptide.

Figure 15
Indication of the TB12.9 immunogenic portions in alignment to the full sequence of TB12.9. Underlined amino acids are different from the TB10.4 peptide.

Preable to examples

It is an established fact that long-term immunological memory resides after termination of a tuberculous infection (Orme, I.M. 1988., Lefford, M.J. et al. 1974.). This memory immunity efficiently protects the host against a secondary infection with *M. tuberculosis* later in life. When an immune host mounts a protective immune response, the specific T-cells responsible for the early recognition of the infected macrophage, stimulates a powerful bactericidal activity through their production of IFN-γ (Rook, G.A.W. 1990., Flesch, I. et al. 1987.). Protective antigens, which are to be incorporated in a future sub-unit vaccine, have in the examples below been sought among the molecular targets of the effector cells responsible for the recall of a protective immune response. This has resulted in the identification of immunodominant antigenic targets for T-cells during the first phase of a protective immune response.

EXAMPLE 1:

Isolation of T-cell stimulating low molecular weight ST-CF antigens

Bacteria. *M. tuberculosis* H37Rv (ATCC 27294) was grown at 37°C on Löwenstein-Jensen medium or in suspension in modified Sauton medium. BCG Copenhagen was obtained as a freeze dried vaccine and were rehydrated with diluted sauton followed by a brief sonication to ensure a disperse suspension.

Production of short-term culture filtrate (ST-CF). ST-CF was produced as described previously (Andersen et al., 1991b). Briefly *M. tuberculosis* (4×10^6 CFU/ml) were incubated in Sauton medium and grown on an orbital shaker for 7 days. The bacteria were removed by filtration and the culture supernatants were passed through sterile filters (0.2 µm) and concentrated on an Amicon YM 3 membrane (Amicon, Danvers, Mass.).

Fractionation of ST-CF by the multi-elution technique. ST-CF (5 mg) was separated in 10-20% SDS-PAGE overnight (11 cm vide centerwell, 0.75 mm gel). After the termination of the electrophoretic run the gel was trimmed for excess gel, and preequilibrated in 3 changes of 2 mM phosphate buffer for 40 min. The multi-elution was performed as described previously (Andersen and Heron, 1993b). Briefly, gels were transferred to the Multi-Eluter™ (KEM-EN-TECH) and electroeluted (40 V) into 2 mM phosphate buffer for 20 min. The polypeptide fractions were aspirated and adjusted to isotonia with concentrated PBS. All fractions were stabilized with 0.5% mice serum and were kept frozen at -80°C until use.

Lymphocyte cultures. Lymphocytes were obtained by preparing single-cell suspensions from spleens as described in Andersen et al., 1991a. Briefly, ST-CF or antigenic fractions were added to microcultures containing 2×10^5 lymphocyte in a volume of 200 µl Rpmi 1640 supplemented with 5×10^5 M 2-mercaptoethanol, penicillin, streptomycin, 1 mM glutamine and 0.5% (vol/vol) fresh mouse serum.

ST-CF was used in the concentration 4 µg/ml while ST-CF fractions were used in 1 µg/ml.

Cellular proliferation was investigated by pulsing the cultures (1 µCi [^{3}H] thymidine/well) after 48 h of incubation, further incubating the plates for 22 hours and finally harvesting and processing

the plates for liquid scintillation counting (Lkb, Beta counter). Culture supernatants were harvested from parallel cultures after 48 hours incubation and used for lymphokine analyses.

Lymphokine analyses. The amount of INF-γ present in culture supernatants and in homogenised organs was quantified by an IFN-γ ELISA kit (Holland Biotechnology, Leiden, the Netherlands). Values below 10 pg were considered negative.

A group of efficiently protected mice was generated by infecting 8-12 weeks old female C57Bl/6j mice bred at Statens Seruminstitut, Copenhagen, Denmark, with 2.5×10^3 *M. tuberculosis* i.v. After 30 days of infection the mice were subjected to 60 days of antibiotic treatment with isoniazid and were then left for 200-240 days to ensure the establishment of resting long-term memory immunity. The mice were then reinfected with 2.5×10^5 *M. tuberculosis* i.v. and the course of infection was compared with that of a corresponding naive group of mice (Fig. 1).

As seen in Fig. 1, *M. tuberculosis* grow rapidly in the spleens of naive mice whereas the infection is controlled within the first few days in memory immune mice. This finding emphasizes that early immunological events occurring during the first days determines the outcome of infection.

Gamma interferon (IFN-γ) is a lymphokine which is involved directly in protective immunity against *M. tuberculosis* (Rook G. A. W., 1990, Flesch I. and Kaufmann S., 1987). To monitor the onset of a protective immune response, the content of IFN-γ in spleen homogenates (4% w/v in PBS) and in serum samples was investigated during the course of infection (Fig. 2). Memory immune mice were found to respond immediately (<24 h) by a marked production of IFN-γ detectable both in spleen and in serum. Naive mice, in contrast, had a 14 days delay before any significant production was evident, a period during which infection rapidly progressed. Immune mice were characterized by an accelerated release of IFN-γ and to determine the molecular targets of this immunological response, spleen lymphocytes were obtained from animals at different time points during the course of infection. The lymphocytes were stimulated *in vitro* with either bacteria, killed with glutaraldehyde and washed with PBS or short-term culture-filtrate (ST-CF) which is a complex mixture of proteins secreted by *M. tuberculosis* during growth (Andersen, P. et al. 1991.) (Fig. 3). The memory immune mice were found to be characterized by an accelerated generation of IFN-γ producing T-cells responding to ST-CF whereas killed

bacteria in contrast were found to elicit only a marginal response at a very late stage of infection.

To map the molecular targets of protective T-cells among the multiple secreted proteins present in ST-CF a screening of ST-CF was performed using the multi-elution technique (Andersen and Heron, 1993). This technique divides complex protein mixtures separated in SDS-PAGE into narrow fractions in a physiological buffer (Fig. 4). These fractions were used to stimulate spleen lymphocytes *in vitro* and the release of IFN-γ was monitored (Fig. 5). The response of long-term memory immune mice (the mice were left for 200-240 days to ensure immunological rest) was compared to the response generated after 4 days of rechallenge infection. This comparison enable the mapping of targets for memory effector T-cells triggered to release IFN-γ during the first phase of a protective immune response. Using this approach it was demonstrated that the targets for these protective T-cells were secreted proteins or fragments of proteins of apparent molecular mass 6-10 and 26-34 kDa (Fig. 5).

To precisely map single molecules within the stimulatory regions the induction of IFN-γ by a panel of narrow overlapping fractions was investigated. This enabled the identification of a 6-8 kDa protein fraction with exceedingly stimulatory capacity (5100-5400 pg IFN-γ units/ml) (Fig. 6). The 6-kDa protein band yielding the highest release of IFN-γ (5390 pg/ml) was recognized by the mAb HYB76-8, whereas the adjacent protein bands were recognized by the mAbs ST-3 and PV-2.

EXAMPLE 2

Cloning of genes expressing low mass culture filtrate antigens

In example 1 it was demonstrated that antigens in the low molecular mass fraction are recognized strongly by cells isolated from memory immune mice. Monoclonal antibodies (mAbs) to these antigens were therefore generated by immunizing with the low mass fraction in RIBI adjuvant (first and second immunization) followed by two injections with the fractions in aluminium hydroxide. Fusion and cloning of the reactive cell lines were done according to standard procedures (Kohler and Milstein 1975). The procedure resulted in the provision of two mAbs: ST-3 directed to a 9 kDa culture filtrate antigen (CFP9) and PV-2 directed to a 7 kDa

antigen (CFP7), when the molecular weight is estimated from migration of the antigens in an SDS-PAGE.

In order to identify the antigens binding to the Mab's, the following experiments were carried out:

The recombinant λgt11 *M. tuberculosis* DNA library constructed by R. Young (Young, R.A. *et al.* 1985) and obtained through the World Health Organization IMMTUB programme (WHO.0032.wibr) was screened for phages expressing gene products which would bind the monoclonal antibodies ST-3 and PV-2.

Approximately 1 x 10^5 pfu of the gene library (containing approximately 25% recombinant phages) were plated on *Eschericia coli* Y1090 (DlacU169, proA$^+$, Dlon, araD139, supF, trpC22::*tn10* [pMC9] ATCC#37197) in soft agar and incubated for 2,5 hours at 42°C.

The plates were overlaid with sheets of nitrocellulose saturated with isopropyl-β-D-thiogalactopyranoside and incubation was continued for 2,5 hours at 37°C. The nitrocellulose was removed and incubated with samples of the monoclonal antibodies in PBS with Tween 20 added to a final concentration of 0.05%. Bound monoclonal antibodies were visualized by horseradish peroxidase-conjugated rabbit anti-mouse immunoglobulins (P260, Dako, Glostrup, DK) and a staining reaction involving 5,5',3,3'-tetramethylbenzidine and H_2O_2.

Positive plaques were recloned and the phages originating from a single plaque were used to lysogenize *E. coli* Y1089 (DlacU169, proA$^+$, Dlon, araD139, strA, hfl150 [chr::*tn10*] [pMC9] ATCC nr. 37196). The resultant lysogenic strains were used to propagate phage particles for DNA extraction. These lysogenic *E. coli* strains have been named:

AA242 (expressing PV-2 reactive polypeptide CFP7) which has been deposited 28 June 1993 with the collection of Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM) under the accession number DSM 8379 and in accordance with the provisions of the Budapest Treaty.

A physical map of the recombinant phages is shown in fig. 7 and the expression of the recombinant gene products is shown in fig. 8.

The PV-2 binding protein appears to be expressed in an unfused version.

Sequencing of the nucleotide sequence encoding the PV-2 binding protein

In order to obtain the nucleotide sequence of the gene encoding the pv-2 binding protein, the approximately 3 kb *M. tuberculosis* derived *EcoR*I - *EcoR*I fragment from AA242 was subcloned in the *EcoR*I site in the pBluescriptSK + (Stratagene) and used to transform *E. coli* XL-1Blue (Stratagene).

The complete DNA sequence of the gene was obtained by the dideoxy chain termination method adapted for supercoiled DNA by use of the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., Cleveland, OH) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. The DNA sequence is shown in SEQ ID NO: 1 (CFP7) as well as in Fig 9. Both strands of the DNA were sequenced.

CFP7

An open reading frame (ORF) encoding a sequence of 96 amino acid residues was identified from an ATG start codon at position 91-93 extending to a TAG stop codon at position 379-381. The deduced amino acid sequence is shown in SEQ ID NO: 2 (and in Fig. 9 where conventional one-letter amino acid codes are used).

CFP7 appear to be expressed in *E. coli* as an unfused version. The nucleotide sequence at position 78-84 is expected to be the Shine Delgarno sequence and the sequences from position 47-50 and 14-19 are expected to be the -10 and -35 regions, respectively:

Subcloning CFP7 in expression vectors

The ORF encoding CFP7 was PCR cloned into the pMST24 (Theisen *et al.*, 1995) expression vector pRVN01.

The PCR amplification was carried out in a thermal reactor (Rapid cycler, Idaho Technology, Idaho) by mixing 10 ng plasmid DNA with the mastermix (0.5 µM of each oligonucleotide primer,

0.25 µM BSA (Stratagene), low salt buffer (20 mM Tris-HCl, pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$ and 0,1% Triton X-100) (Stratagene), 0.25 mM of each deoxynucleoside triphosphate and 0.5 U Taq Plus Long DNA polymerase (Stratagene)). Final volume was 10 µl (all concentrations given are concentrations in the final volume). Predenaturation was carried out at 94°C for 30 s. 30 cycles of the following was performed; Denaturation at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 1 min.

The oligonucleotide primers were synthesised automatically on a DNA synthesizer (Applied Biosystems, Forster City, Ca, ABI-391, PCR-mode), deblocked, and purified by ethanol precipitation.

The *cfp7* oligonucleotides (TABLE 1) were synthesised on the basis of the nucleotide sequence from the CFP7 sequence (Fig. 9). The oligonucleotides were engineered to include an *Sma*I restriction enzyme site at the 5' end and a *Bam*HI restriction enzyme site at the 3' end for directed subcloning.

CFP7

By the use of PCR a *Sma*I site was engineered immediately 5' of the first codon of the ORF of 291 bp, encoding the *cfp7* gene, so that only the coding region would be expressed, and a *Bam*HI site was incorporated right after the stop codon at the 3' end. The 291 bp PCR fragment was cleaved by *Sma*I and *Bam*HI, purified from an agarose gel and subcloned into the *Sma*I - *Bam*HI sites of the pMST24 expression vector. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue (pRVN01).

Purification of recombinant CFP7

The ORF was fused N-terminally to the $(His)_6$-tag (cf. EP-A-0 282 242). Recombinant antigen was prepared as follows: Briefly, a single colony of *E. coli* harbouring either the pRVN01 or the pRVN02 plasmid, was inoculated into Luria-Bertani broth containing 100 µg/ml ampicillin and 12.5 µg/ml tetracycline and grown at 37°C to OD_{600nm} = 0.5. IPTG (isopropyl-β-D-thiogalactoside) was then added to a final concentration of 2 mM (expression was regulated either by the strong IPTG inducible P_{tac} or the T5 promoter) and growth was continued for further 2 hours. The cells were harvested by centrifugation at 4,200 x g at 4°C for 8 min. The

pelleted bacteria were stored overnight at -20°C. The pellet was resuspended in BC 40/100 buffer (20 mM Tris-HCl pH 7.9, 20% glycerol, 100 mM KCl, 40 mM Imidazole) and cells were broken by sonication (5 times for 30 s with intervals of 30 s) at 4°C. followed by centrifugation at 12,000 x g for 30 min at 4°C, the supernatant (crude extract) was used for purification of the recombinant antigens.

The Histidine fusion protein (His-rCFP7) was purified from the crude extract by affinity chromatography on a Ni^{2+}-NTA column from QIAGEN with a volume of 100 ml. His-rCFP7 binds to Ni^{2+}. After extensive washes of the column in BC 40/100 buffer, the fusion protein was eluted with a BC 1000/100 buffer containing 100 mM imidazole, 20 mM Tris pH 7.9, 20% glycerol and 1 M KCl. subsequently, the purified products were dialysed extensively against 10 mM Tris pH 8.0. His-rCFP7 was then separated from contaminants by fast protein liquid chromatography (FPLC) over an anion-exchange column (Mono Q, Pharmacia, Sweden) in 10 mM Tris pH 8.0 with a linear gradient of NaCl from 0 to 1 M. Aliquots of the fractions were analyzed by 10%-20% gradient sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Fractions containing purified His-rCFP7 was pooled.

TABLE 1. Sequence of the *cfp7* oligonucleotides[a].

Orientation and oligonucleotide	Sequences (5' → 3')	Position[b] (nucleotide)
Sense		
pvR3	GCAACACCCGGGATGTCGCAAATCATG (SEQ ID NO: 43)	91-105 (SEQ ID NO: 1)
Antisense		
pvF4	CTACTAAGCTTGGATCCCTAGCCG-CCCCATTTGGCGG (SEQ ID NO: 45)	381-362 (SEQ ID NO: 1)

[a] The *cfp7* oligonucleotides were based on the nucleotide sequence shown in Fig. 9 (SEQ ID NO: 1).
Nucleotides underlined are not contained in the nucleotide sequence of *cfp7*.
[b] The positions referred to are of the non-underlined part of the primers and correspond to the nucleotide sequence shown in Fig. 9 and Fig. 10, respectively.

EXAMPLE 3

Identification of antigens which are not expressed in BCG strains.

In an effort to control the treat of TB, attenuated bacillus Calmette-Guérin (BCG) has been used as a live attenuated vaccine. BCG is an attenuated derivative of a virulent *Mycobacterium bovis*. The original BCG from the Pasteur Institute in Paris, France was developed from 1908 to 1921 by 231 passages in liquid culture and has never been shown to revert to virulence in animals, indicating that the attenuating mutation(s) in BCG are stable deletions and/or multiple mutations which do not readily revert. While physiological differences between BCG and *M. tuberculosis* and *M. bovis* has been noted, the attenuating mutations which arose during serial passage of the original BCG strain has been unknown until recently. The first mutations described are the loss of the gene encoding MPB64 in some BCG strains (Li et al., 1993, Oettinger and Andersen, 1994) and the gene encoding ESAT-6 in all BCG strain tested (Harboe et al., 1996), later 3 large deletions in BCG have been identified (Mahairas et al., 1996). The region named RD1

includes the gene encoding ESAT-6 and an other (RD2) the gene encoding MPT64. Both antigens have been shown to have diagnostic potential and ESAT-6 has been shown to have properties as a vaccine candidate (cf. PCT/DK94/00273 and PCT/DK94/00270). In order to find new *M. tuberculosis* specific diagnostic antigens as well as antigens for a new vaccine against TB, the RD1 region (17.499 bp) of *M. tuberculosis* H37Rv has been analyzed for Open Reading Frames (ORF). ORFs with a minimum length of 96 bp have been predicted using the algorithm described by Borodovsky and McIninch (1993), in total 27 ORFs have been predicted, 20 of these have possible diagnostic and/or vaccine potential, as they are deleted from all known BCG strains. The predicted ORFs include ESAT-6 (RD1-ORF7) and CFP10 (RD1-ORF6) described previously (Sørensen et al., 1995), as a positive control for the ability of the algorithm. In the present is described the potential of 7 of the predicted antigens for diagnosis of TB as well as potential as candidates for a new vaccine against TB.

Seven open reading frames (ORF) from the 17,499kb RD1 region (Accession no. U34848) with possible diagnostic and vaccine potential have been identified and cloned.

Identification of the ORF's *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a,* and *rd1-orf9b.*

The nucleotide sequence of *rd1-orf2* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 71. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 72.

The nucleotide sequence of *rd1-orf3* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 87. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 88.

The nucleotide sequence of *rd1-orf4* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 89. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 90.

The nucleotide sequence of *rd1-orf5* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 91. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 92.

The nucleotide sequence of *rd1-orf8* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 67. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 68.

The nucleotide sequence of *rd1-orf9a* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 93. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 94.

The nucleotide sequence of *rd1-orf9b* from *M. tuberculosis* H37Rv is set forth in SEQ ID NO: 69. The deduced amino acid sequence of RD1-ORF2 is set forth in SEQ ID NO: 70.

The DNA sequence *rd1-orf2* (SEQ ID NO: 71) contained an open reading frame starting with an ATG codon at position 889 - 891 and ending with a termination codon (TAA) at position 2662 - 2664 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 72) contains 591 residues corresponding to a molecular weight of 64,525.

The DNA sequence *rd1-orf3* (SEQ ID NO: 87) contained an open reading frame starting with an ATG codon at position 2807 - 2809 and ending with a termination codon (TAA) at position 3101 - 3103 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 88) contains 98 residues corresponding to a molecular weight of 9,799.

The DNA sequence *rd1-orf4* (SEQ ID NO: 89) contained an open reading frame starting with a GTG codon at position 4014 - 4012 and ending with a termination codon (TAG) at position 3597 - 3595 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 90) contains 139 residues corresponding to a molecular weight of 14,210.

The DNA sequence *rd1-orf5* (SEQ ID NO: 91) contained an open reading frame starting with a GTG codon at position 3128 - 3130 and ending with a termination codon (TGA) at position 4241 - 4243 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 92) contains 371 residues corresponding to a molecular weight of 37,647.

The DNA sequence *rd1-orf8* (SEQ ID NO: 67) contained an open reading frame starting with a GTG codon at position 5502 - 5500 and ending with a termination codon (TAG) at position 5084 - 5082 (position numbers referring to the location in RD1), and the deduced amino acid sequence (SEQ ID NO: 68) contains 139 residues with a molecular weight of 11,737.

The DNA sequence *rd1-orf9a* (SEQ ID NO: 93) contained an open reading frame starting with a GTG codon at position 6146 - 6148 and ending with a termination codon (TAA) at position 7070

- 7072 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 94) contains 308 residues corresponding to a molecular weight of 33,453.

The DNA sequence *rd1-orf9b* (SEQ ID NO: 69) contained an open reading frame starting with an ATG codon at position 5072 - 5074 and ending with a termination codon (TAA) at position 7070 - 7072 (position numbers referring to the location in RD1). The deduced amino acid sequence (SEQ ID NO: 70) contains 666 residues corresponding to a molecular weight of 70,650.

Cloning of the ORF's *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a,* and *rd1-orf9b.*

The ORF's *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* were PCR cloned in the pMST24 (Theisen et al., 1995) (*rd1-orf3*) or the pQE32 (QIAGEN) (*rd1-orf2, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b*) expression vector. Preparation of oligonucleotides and PCR amplification of the *rd1-orf* encoding genes, was carried out as described in example 2. Chromosomal DNA from *M. tuberculosis* H37Rv was used as template in the PCR reactions. Oligonucleotides were synthesized on the basis of the nucleotide sequence from the RD1 region (Accession no. U34848). The oligonucleotide primers were engineered to include an restriction enzyme site at the 5' end and at the 3' end by which a later subcloning was possible. Primers are listed in TABLE 2.

rd1-orf2. A *Bam*HI site was engineered immediately 5' of the first codon of *rd1-orf2*, and a *Hind*III site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf2* was subcloned in pQE32, giving pTO96.

rd1-orf3. A *Sma*I site was engineered immediately 5' of the first codon of *rd1-orf3*, and a *Nco*I site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf3* was subcloned in pMST24, giving pTO87.

rd1-orf4. A *Bam*HI site was engineered immediately 5' of the first codon of *rd1-orf4*, and a *Hind*III site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf4* was subcloned in pQE32, giving pTO89.

rd1-orf5. A *Bam*HI site was engineered immediately 5' of the first codon of *rd1-orf5*, and a *Hind*III site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf5* was subcloned in pQE32, giving pTO88.

rd1-orf8. A *Bam*HI site was engineered immediately 5' of the first codon of *rd1-orf8*, and a *Nco*I site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf8* was subcloned in pMST24, giving pTO98.

rd1-orf9a. A *Bam*HI site was engineered immediately 5' of the first codon of *rd1-orf9a*, and a *Hind*III site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf9a* was subcloned in pQE32, giving pTO91.

rd1-orf9b. A *Sca*I site was engineered immediately 5' of the first codon of *rd1-orf9b*, and a *Hind* III site was incorporated right after the stop codon at the 3' end. The gene *rd1-orf9b* was subcloned in pQE32, giving pTO90.

The PCR fragments were digested with the suitable restriction enzymes, purified from an agarose gel and cloned into either pMST24 or pQE-32. The seven constructs were used to transform the *E. coli* XL1-Blue. Endpoints of the gene fusions were determined by the dideoxy chain termination method. Both strands of the DNA were sequenced.

Purification of recombinant RD1-ORF2, RD1-ORF3, RD1-ORF4, RD1-ORF5, RD1-ORF8, RD1-ORF9a and RD1-ORF9b.

The rRD1-ORFs were fused N-terminally to the $(His)_6$ -tag. Recombinant antigen was prepared as described in example 2 (with the exception that pTO91 was expressed at 30°C and not at 37°C), using a single colony of *E. coli* harbouring either the pTO87, pTO88, pTO89, pTO90, pTO91, pTO96 or pTO98 for inoculation. Purification of recombinant antigen by Ni^{2+} affinity chromatography was also carried out as described in example 2. Fractions containing purified His-rRD1-ORF2, His-rRD1-ORF3 His-rRD1-ORF4, His-rRD1-ORF5, His-rRD1-ORF8, His-rRD1-ORF9a or His-rRD1-ORF9b were pooled. The His-rRD1-ORF's were extensively dialysed against 10 mM Tris/HCl, pH 8.5, 3 M urea followed by an additional purification step performed on an anion exchange column (Mono Q) using fast protein liquid chromatography (FPLC) (Pharmacia, Uppsala, Sweden). The purification was carried out in 10 mM Tris/HCl, pH 8.5, 3 M

urea and protein was eluted by a linear gradient of NaCl from 0 to 1 M. Fractions containing the His-rRD1-ORF's were pooled and subsequently dialysed extensively against 25 mM Hepes, pH 8.0 before use.

Table 2. Sequence of the *rd1-orf's* oligonucleotides[a].

Orientation and oligo-nucleotide	Sequences (5'→ 3')	Position (nt)
Sense		
RD1-ORF2f	CTGGGGATCCGCATGACTGCTGAACCG	886 - 903
RD1-ORF3f	CTTCCCGGGATGGAAAAAATGTCAC	2807 - 2822
RD1-ORF4f	GTAGGATCCTAGGAGACATCAGCGGC	4028 - 4015
RD1-ORF5f	CTGGGGATCCGCGTGATCACCAT-GCTGTGG	3028 - 3045
RD1-ORF8f	CTCGGATCCTGTGGGTGCAGGTCCGGC GATGGGC	5502 - 5479
RD1-ORF9af	GTGATGTGAGCTCAGGTGAAGAA-GGTGAAG	6144 - 6160
RD1-ORF9bf	GTGATGTGAGCTCCTATGGCGGCCGAC-TACGAC	5072 - 5089
Antisense		
RD1-ORF2r	TGCAAGCTTTTAACCGGCGCTTGGGGGT GC	2664 - 2644
RD1-ORF3r	GATGCCATGGTTAGGCGAAGACGC-CGGC	3103 - 3086
RD1-ORF4r	CGATCTAAGCTTGGCAATGGAGGTCTA	3582 - 3597
RD1-ORF5r	TGCAAGCTTTCACCAGTCGTCCT-CTTCGTC	4243 - 4223
RD1-ORF8r	CTCCCATGGCTACGACAAGCTCTTC-CGGCCGC	5083 - 5105
RD1-ORF9a/br	CGATCTAAGCTTTCAACGACGTCCAGCC	7073 - 7056

[a] The oligonucleotides were constructed from the Accession number U34484 nucleotide sequence (Mahairas et al., 1996). Nucleotides (nt) underlined are not contained in the nucleotide sequence of RD1-ORF's. The positions correspond to the nucleotide sequence of Accession number U34484.

The nucleotide sequences of *rd1-orf2*, *rd1-orf3*, *rd1-orf4*, *rd1-orf5*, *rd1-orf8*, *rd1-orf9a*, and *rd1-orf9b* from *M. tuberculosis* H37Rv are set forth in SEQ ID NO: 71, 87, 89, 91, 67, 93, and 69, respectively. The deduced amino acid sequences of *rd1-orf2*, *rd1-orf3*, *rd1-orf4* *rd1-orf5*, *rd1-orf8*, *rd1-orf9a*, and *rd1-orf9b* are set forth in SEQ ID NO: 72, 88, 90, 92, 68, 94, and 70, respectively.

EXAMPLE 4

Cloning of the genes expressing 17-30 kDa antigens from ST-CF

Isolation of CFP17, CFP20, CFP21, CFP22, CFP25, and CFP28

ST-CF was precipitated with ammonium sulphate at 80% saturation. The precipitated proteins were removed by centrifugation and after resuspension washed with 8 M urea. CHAPS and glycerol were added to a final concentration of 0.5% (w/v) and 5% (v/v) respectively and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad). The Rotofor Cell had been equilibrated with an 8 M urea buffer containing 0.5% (w/v) CHAPS, 5% (v/v) glycerol, 3% (v/v) Biolyt 3/5 and 1% (v/v) Biolyt 4/6 (BioRad). Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed on silver-stained 10-20% SDS-PAGE. Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml. An equal volume of SDS containing sample buffer was added and the protein solution boiled for 5 min before further separation on a Prep Cell (BioRad) in a matrix of 16% polyacrylamide under an electrical gradient. Fractions containing pure proteins with an molecular mass from 17-30 kDa were collected.

Isolation of CFP29

Anti-CFP29, reacting with CFP29 was generated by immunization of BALB/c mice with crushed gel pieces in RIBI adjuvant (first and second immunization) or aluminium hydroxide (third immunization and boosting) with two week intervals. SDS-PAGE gel pieces containing 2-5 µg of CFP29 were used for each immunization. Mice were boosted with antigen 3 days before removal of the spleen. Generation of a monoclonal cell line producing antibodies against CFP29 was obtained essentially as described by Köhler and Milstein (1975). Screening of supernatants from growing clones was carried out by immunoblotting of nitrocellulose strips containing ST-CF

separated by SDS-PAGE. Each strip contained approximately 50 µg of ST-CF. The antibody class of anti-CFP29 was identified as IgM by the mouse monoclonal antibody isotyping kit, RPN29 (Amersham) according to the manufacturer's instructions.

CFP29 was purified by the following method: ST-CF was concentrated 10 fold by ultrafiltration, and ammonium sulphate precipitation in the 45 to 55% saturation range was performed. The pellet was redissolved in 50 mM sodium phosphate, 1.5 M ammonium sulphate, pH 8.5, and subjected to thiophilic adsorption chromatography (Porath *et al.*, 1985) on an Affi-T gel column (Kem-En-Tec). Protein was eluted by a linear 1.5 to 0 M gradient of ammonium sulphate and fractions collected in the range 0.44 to 0.31 M ammonium sulphate were identified as CFP29 containing fractions in Western blot experiments with mAb Anti-CFP29. These fractions were pooled and anion exchange chromatography was performed on a Mono Q HR 5/5 column connected to an FPLC system (Pharmacia). The column was equilibrated with 10 mM Tris-HCl, pH 8.5 and the elution was performed with a linear gradient from 0 to 500 mM NaCl. From 400 to 500 mM sodium chloride, rather pure CFP29 was eluted. As a final purification step the Mono Q fractions containing CFP29 were loaded on a 12.5% SDS-PAGE gel and pure CFP29 was obtained by the multi-elution technique (Andersen and Heron, 1993).

N-terminal sequencing and amino acid analysis

CFP17, CFP20, CFP21, CFP22, CFP25, and CFP28 were washed with water on a Centricon concentrator (Amicon) with cutoff at 10 kDa and then applied to a ProSpin concentrator (Applied Biosystems) where the proteins were collected on a PVDF membrane. The membrane was washed 5 times with 20% methanol before sequencing on a Procise sequencer (Applied Biosystems).

CFP29 containing fractions were blotted to PVDF membrane after tricine SDS-PAGE (Ploug *et al.*, 1989). The relevant bands were excised and subjected to amino acid analysis (Barkholt and Jensen, 1989) and N-terminal sequence analysis on a Procise sequencer (Applied Biosystems).

The following N-terminal sequences were obtained:

For CFP17: A/S E L D A P A Q A G T E X A V (SEQ ID NO: 17)

For CFP20: A Q I T L R G N A I N T V G E (SEQ ID NO: 18)

For CFP21: D P X S D I A V V F A R G T H (SEQ ID NO: 19)
For CFP22: T N S P L A T A T A T L H T N (SEQ ID NO: 20)
For CFP25: A X P D A E V V F A R G R F E (SEQ ID NO: 21)
For CFP28: X I/V Q K S L E L I V/T V/F T A D/Q E (SEQ ID NO: 22)
For CFP29: M N N L Y R D L A P V T E A A W A E I (SEQ ID NO: 23)

"X" denotes an amino acid which could not be determined by the sequencing method used, whereas a "/" between two amino acids denotes that the sequencing method could not determine which of the two amino acids is the one actually present.

Cloning the gene encoding CFP29

The N-terminal sequence of CFP29 was used for a homology search in the EMBL database using the TFASTA program of the Genetics Computer Group sequence analysis software package. The search identified a protein, Linocin M18, from *Brevibacterium linens* that shares 74% identity with the 19 N-terminal amino acids of CFP29.

Based on this identity between the N-terminal sequence of CFP29 and the sequence of the Linocin M18 protein from *Brevibacterium linens*, a set of degenerated primers were constructed for PCR cloning of the *M. tuberculosis* gene encoding CFP29. PCR reactions were containing 10 ng of *M. tuberculosis* chromosomal DNA in 1 × low salt Taq+ buffer from Stratagene supplemented with 250 µM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 µl reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 15 sec., 55°C for 15 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.

An approx. 300 bp fragment was obtained using primers with the sequences:

1: 5'-CCCGGCTCGAGAACCTSTACCGCGACCTSGCSCC (SEQ ID NO: 24)
2: 5'-GGGCCGGATCCGASGCSGCGTCCTTSACSGGYTGCCA (SEQ ID NO: 25)
-where S = G/C and Y = T/C

The fragment was excised from a 1% agarose gel, purified by Spin-X spinn columns (Costar), cloned into pBluescript SK II+ - T vector (Stratagene) and finally sequenced with the Sequenase kit from United States Biochemical.

The first 150 bp of this sequence was used for a homology search using the Blast program of the Sanger *Mycobacterium tuberculosis* database:

(http//www.sanger.ac.uk/projects/M-tuberculosis/blast_server).

This program identified a *Mycobacterium tuberculosis* sequence on cosmid cy444 in the database that is nearly 100% identical to the 150 bp sequence of the CFP29 protein. The sequence is contained within a 795 bp open reading frame of which the 5' end translates into a sequence that is 100% identical to the N-terminally sequenced 19 amino acids of the purified CFP29 protein.

Finally, the 795 bp open reading frame was PCR cloned under the same PCR conditions as described above using the primers:

3: 5'-GGAAGCCCCATATGAACAATCTCTACCG (SEQ ID NO: 26)

4: 5'-CGCGCTCAGCCCTTAGTGACTGAGCGCGACCG (SEQ ID NO: 27)

The resulting DNA fragments were purified from agarose gels as described above sequenced with primer 3 and 4 in addition to the following primers:

5: 5'-GGACGTTCAAGCGACACATCGCCG-3' (SEQ ID NO: 115)

6: 5'-CAGCACGAACGCGCCGTCGATGGC-3' (SEQ ID NO: 116)

Three independent cloned were sequenced. All three clones were in 100% agreement with the sequence on cosmid cy444.

All other DNA manipulations were done according to Maniatis *et al.* (1989).

All enzymes other than Taq polymerase were from New England Biolabs.

Homology searches in the Sanger database

For CFP17, CFP20, CFP21, CFP22, CFP25, and CFP28 the N-terminal amino acid sequence from each of the proteins were used for a homology search using the blast program of the Sanger *Mycobacterium tuberculosis* database:

http://www.sanger.ac.uk/pathogens/TB-blast-server.html.

For CFP29 the first 150 bp of the DNA sequence was used for the search. Furthermore, the EMBL database was searched for proteins with homology to CFP29.

Thereby, the following information were obtained:

CFP17

Of the 14 determined amino acids in CFP17 a 93% identical sequence was found with MTCY1A11.16c. The difference between the two sequences is in the first amino acid: It is an A or an S in the N-terminal determined sequenced and a S in MTCY1A11. From the N-terminal sequencing it was not possible to determine amino acid number 13.

Within the open reading frame the translated protein is 162 amino acids long. The N-terminal of the protein purified from culture filtrate starts at amino acid 31 in agreement with the presence of a signal sequence that has been cleaved off. This gives a length of the mature protein of 132 amino acids, which corresponds to a theoretical molecular mass of 13833 Da and a theoretical pI of 4.4. The observed mass in SDS-PAGE is 17 kDa.

CFP20

A sequence 100% identical to the 15 determined amino acids of CFP20 was found on the translated cosmid cscy09F9. A stop codon is found at amino acid 166 from the amino acid M at position 1. This gives a predicted length of 165 amino acids, which corresponds to a theoretical molecular mass of 16897 Da and a pI of 4.2. The observed molecular weight in a SDS-PAGE is 20 kDa.

Searching the GenEMBL database using the TFASTA algorithm (Pearson and Lipman, 1988) revealed a number of proteins with homology to the predicted 164 amino acids long translated protein.

The highest homology, 51.5% identity in a 163 amino acid overlap, was found to a Haemophilus influenza Rd toxR reg. (HIHI0751).

CFP21

A sequence 100% identical to the 14 determined amino acids of CFP21 was found at MTCY39. From the N-terminal sequencing it was not possible to determine amino acid number 3; this amino acid is a C in MTCY39. The amino acid C can not be detected on a Sequencer which is probably the explanation of this difference.

Within the open reading frame the translated protein is 217 amino acids long. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 33 in agreement with the presence of a signal sequence that has been cleaved off. This gives a length of the mature protein of 185 amino acids, which corresponds to a theoretical molecular weigh at 18657 Da, and a theoretical pI at 4,6. The observed weight in a SDS-PAGE is 21 kDa.

In a 193 amino acids overlap the protein has 32,6% identity to a cutinase precursor with a length of 209 amino acids (CUTI_ALTBR P41744).

A comparison of the 14 N-terminal determined amino acids with the translated region (RD2) deleted in *M. bovis* BCG revealed a 100% identical sequence (mb3484) (Mahairas *et al.* (1996)).

CFP22

A sequence 100% identical to the 15 determined amino acids of CFP22 was found at MTCY10H4. Within the open reading frame the translated protein is 182 amino acids long. The N-terminal sequence of the protein purified from culture filtrate starts at amino acid 8 and therefore the length of the protein occurring in *M. tuberculosis* culture filtrate is 175 amino acids.

This gives a theoretical molecular weigh at 18517 Da and a pI at 6.8. The observed weight in a SDS-PAGE is 22 kDa.

In an 182 amino acids overlap the translated protein has 90,1% identity with E235739; a peptidyl-prolyl cis-trans isomerase.

CFP25

A sequence 93% identical to the 15 determined amino acids was found on the cosmid MTCY339.08c. The one amino acid that differs between the two sequences is a C in MTCY339.08c and a X from the N-terminal sequence data. On a Sequencer a C can not be detected which is a probable explanation for this difference.

The N-terminally determined sequence from the protein purified from culture filtrate begins at amino acid 33 in agreement with the presence of a signal sequence that has been cleaved off. This gives a length of the mature protein of 187 amino acids, which corresponds to a theoretical molecular weigh at 19665 Da, and a theoretical pI at 4.9. The observed weight in a SDS-PAGE is 25 kDa.

In a 217 amino acids overlap the protein has 42.9% identity to CFP21 (MTCY39.35).

CFP28

No homology was found when using the 10 determined amino acid residues 2-8, 11, 12, and 14 of SEQ ID NO: 22 in the database search.

CFP29

Sanger database searching: A sequence nearly 100% identical to the 150 bp sequence of the CFP29 protein was found on cosmid cy444. The sequence is contained within a 795 bp open reading frame of which the 5' end translates into a sequence that is 100% identical to the N-terminally sequenced 19 amino acids of the purified CFP29 protein. The open reading frame encodes a 265 amino acid protein.

The amino acid analysis performed on the purified protein further confirmed the identity of CFP29 with the protein encoded in open reading frame on cosmid 444.

EMBL database searching: The open reading frame encodes a 265 amino acid protein that is 58% identical and 74% similar to the Linocin M18 protein (61% identity on DNA level). This is a 28.6 kDa protein with bacteriocin activity (Valdés-Stauber and Scherer, 1994; Valdés-Stauber and Scherer, 1996). The two proteins have the same length (except for 1 amino acid) and share the same theoretical physicochemical properties. We therefore suggest that CFP29 is a mycobacterial homolog to the *Brevibacterium linens* Linocin M18 protein.

The amino acid sequences of the purified antigens as picked from the Sanger database are shown in the following list. The amino acids determined by N-terminal sequencing are marked with bold.

CFP17 (SEQ ID NO: 6):

1 MTDMNPDIEK DQTSDEVTVE TTSVFRADFL **SELDAPAQAG TESAV**SGVEG
51 LPPGSALLVV KRGPNAGSRF LLDQAITSAG RHPDSDIFLD DVTVSRRHAE
101 FRLENNEFNV VDVGSLNGTY VNREPVDSAV LANGDEVQIG KFRLVFLTGP
151 KQGEDDGSTG GP

CFP20 (SEQ ID NO: 8):

1 M**AQITLRGNA INTVGE**LPAV GSPAPAFTLT GGDLGVISSD QFRGKSVLLN
51 IFPSVDTPVC ATSVRTFDER AAASGATVLC VSKDLPFAQK RFCGAEGTEN
101 VMPASAFRDS FGEDYGVTIA DGPMAGLLAR AIVVIGADGN VAYTELVPEI
151 AQEPNYEAAL AALGA

CFP21 (SEQ ID NO: 10):

1 MTPRSLVRIV GVVVATTLAL VSAPAGGRAA HA**DPCSDIAV**
41 **VFARGTH**QAS GLGDVGEAFV DSLTSQVGGR SIGVYAVNYP ASDDYRASAS
91 NGSDDASAHI QRTVASCPNT RIVLGGYSQG ATVIDLSTSA MPPAVADHVA
141 AVALFGEPSS GFSSMLWGGG SLPTIGPLYS SKTINLCAPD DPICTGGGNI

191 MAHVSYVQSG MTSQAATFAA NRLDHAG

CFP22 (SEQ ID NO: 12):

1 MADCDSV**TNS PLATATATLH TN**RGDIKIAL FGNHAPKTVA NFVGLAQGTK
51 DYSTQNASGG PSGPFYDGAV FHRVIQGFMI QGGDPTGTGR GGPGYKFADE
101 FHPELQFDKP YLLAMANAGP GTNGSQFFIT VGKTPHLNRR HTIFGEVIDA
151 ESQRVVEAIS KTATDGNDRP TDPVVIESIT IS

CFP25 (SEQ ID NO: 14):

1 MGAAAAMLAA VLLLTPITVP AGYPGAVAPA TA**ACPDAEVV FARGRFE**PPG
51 IGTVGNAFVS ALRSKVNKNV GVYAVKYPAD NQIDVGANDM SAHIQSMANS
101 CPNTRLVPGG YSLGAAVTDV VLAVPTQMWG FTNPLPPGSD EHIAAVALFG
151 NGSQWVGPIT NFSPAYNDRT IELCHGDDPV CHPADPNTWE ANWPQHLAGA
201 YVSSGMVNQA ADFVAGKLQ

CFP29 (SEQ ID NO: 16):

1 **MNNLYRDLAP VTEAAWAEI**E LEAARTFKRH IAGRRVVDVS DPGGPVTAAV
51 STGRLIDVKA PTNGVIAHLR ASKPLVRLRV PFTLSRNEID DVERGSKDSD
101 WEPVKEAAKK LAFVEDRTIF EGYSAASIEG IRSASSNPAL TLPEDPREIP
151 DVISQALSEL RLAGVDGPYS VLLSADVYTK VSETSDHGYP IREHLNRLVD
201 GDIIWAPAID GAFVLTTRGG DFDLQLGTDV AIGYASHDTD TVRLYLQETL
251 TFLCYTAEAS VALSH

For all six proteins the molecular weights predicted from the sequences are in agreement with the molecular weights observed on SDS-PAGE.

Cloning of the genes encoding CFP17, CFP20, CFP21, CFP22 and CFP25.

The genes encoding CFP17, CFP20, CFP21, CFP22 and CFP25 were all cloned into the expression vector pMCT6, by PCR amplification with gene specific primers, for recombinant expression in *E. coli* of the proteins.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1x low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Tag+ DNA polymerase (Stratagene) in 10 µl reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles according to the following program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluescript SK II+ - T vector (Stratagene). Plasmid DNA was thereafter prepared from clones harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMCT6 in frame with 8 histidine residues which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

For cloning of the individual antigens, the following gene specific primers were used:

<u>CFP17:</u> Primers used for cloning of cfp17:

OPBR-51: ACAGATCTGTGACGGACATGAACCCG (SEQ ID NO: 117)
OPBR-52: TTTTCCATGGTCACGGGCCCCCGGTACT (SEQ ID NO: 118)

OPBR-51 and OPBR-52 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

<u>CFP20:</u> Primers used for cloning of cfp20:

OPBR-53: ACAGATCTGTGCCCATGGCACAGATA (SEQ ID NO: 119)
OPBR-54: TTTAAGCTTCTAGGCGCCCAGCGCGGC (SEQ ID NO: 120)

OPBR-53 and OPBR-54 create BglII and HinDIII sites, respectively, used for the cloning in pMCT6.

CFP21: Primers used for cloning of cfp21:

OPBR-55: ACAGATCTGCGCATGCGGATCCGTGT (SEQ ID NO: 121)
OPBR-56: TTTTCCATGGTCATCCGGCGTGATCGAG (SEQ ID NO: 122)

OPBR-55 and OPBR-56 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

CFP22: Primers used for cloning of cfp22:

OPBR-57: ACAGATCTGTAATGGCAGACTGTGAT (SEQ ID NO: 123)
OPBR-58: TTTTCCATGGTCAGGAGATGGTGATCGA (SEQ ID NO: 124)

OPBR-57 and OPBR-58 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

CFP25: Primers used for cloning of cfp25:

OPBR-59: ACAGATCTGCCGGCTACCCCGGTGCC (SEQ ID NO: 125)
OPBR-60: TTTTCCATGGCTATTGCAGCTTTCCGGC (SEQ ID NO: 126)

OPBR-59 and OPBR-60 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

Expression/purification of recombinant CFP17, CFP20, CFP21, CFP22 and CFP25 proteins.

Expression and metal affinity purification of recombinant proteins was undertaken essentially as described by the manufacturers. For each protein, 1 l LB-media containing 100 µg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmids. Cultures were shaken at 37 °C until they reached a density of OD_{600} = 0.4 -

0.6. IPTG was hereafter added to a final concentration of 1 mM and the cultures were further incubated 4 - 16 hours. Cells were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.
After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analyzed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content were determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 5

Identification of CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19B, CFP22A, CFP23A, CFP23B, CFP25A, CFP27, CFP30A, CWP32 and CFP50.

Identification of CFP16 and CFP19B.

ST-CF was precipitated with ammonium sulphate at 80% saturation. The precipitated proteins were removed by centrifugation and after resuspension washed with 8 M urea. CHAPS and glycerol were added to a final concentration of 0.5 % (w/v) and 5 % (v/v) respectively and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad). The Rotofor Cell had been equilibrated with a 8M urea buffer containing 0.5 % (w/v) CHAPS, 5% (v/v) glycerol, 3% (v/v) Biolyt 3/5 and 1% (v/v) Biolyt 4/6 (BioRad). Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed on silver-stained 10-20% SDS-PAGE. Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml. An equal

volume of SDS containing sample buffer was added and the protein solution boiled for 5 min before further separation on a Prep Cell (BioRad) in a matrix of 16% polyacrylamide under an electrical gradient. Fractions containing well separated bands in SDS-PAGE were selected for N-terminal sequencing after transfer to PVDF membrane.

Isolation of CFP8A, CFP8B, CFP19, CFP23A, and CFP23B.

ST-CF was precipitated with ammonium sulphate at 80% saturation and redissolved in PBS, pH 7.4, and dialysed 3 times against 25mM Piperazin-HCl, pH 5.5, and subjected to chromatofocusing on a matrix of PBE 94 (Pharmacia) in a column connected to an FPLC system (Pharmacia). The column was equilibrated with 25 mM Piperazin-HCl, pH 5.5, and the elution was performed with 10% PB74-HCl, pH 4.0 (Pharmacia). Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml and separated on a Prepcell as described above.

Identification of CFP22A

ST-CF was concentrated approximately 10 fold by ultrafiltration and proteins were precipitated at 80 % saturation, redissolved in PBS, pH 7.4, and dialysed 3 times against PBS, pH 7.4. 5.1 ml of the dialysed ST-CF was treated with RNase (0.2 mg/ml, QUIAGEN) and DNase (0.2 mg/ml, Boehringer Mannheim) for 6 h and placed on top of 6.4 ml of 48 % (w/v) sucrose in PBS, pH 7.4, in Sorvall tubes (Ultracrimp 03987, DuPont Medical Products) and ultracentrifuged for 20 h at 257,300 × g_{max}, 10°C. The pellet was redissolved in 200 µl of 25 mM Tris-192 mM glycine, 0.1 % SDS, pH 8.3.

Identification of CFP7A, CFP25A, CFP27, CFP30A and CFP50

For CFP27, CFP30A and CFP50 ST-CF was concentrated approximately 10 fold by ultrafiltration and ammonium sulphate precipitation in the 45 to 55 % saturation range was performed. Proteins were redissolved in 50 mM sodium phosphate, 1.5 M ammonium sulphate, pH 8.5, and subjected to thiophilic adsorption chromatography on an Affi-T gel column (Kem-En-Tec). Proteins were eluted by a 1.5 to 0 M decreasing gradient of ammonium sulphate. Fractions with similar band patterns in SDS-PAGE were pooled and anion exchange chroma-

tography was performed on a Mono Q HR 5/5 column connected to an FPLC system (Pharmacia). The column was equilibrated with 10 mM Tris-HCl, pH 8.5, and the elution was performed with a gradient of NaCl from 0 to 1 M. Fractions containing well separated bands in SDS-PAGE were selected.

CFP7A and CFP25A were obtained as described above except for the following modification: ST-CF was concentrated approximately 10 fold by ultrafiltration and proteins were precipitated at 80 % saturation, redissolved in PBS, pH 7.4, and dialysed 3 times against PBS, pH 7.4. Ammonium sulphate was added to a concentration of 1.5 M, and ST-CF proteins were loaded on an Affi T-gel column. Elution from the Affi T-gel column and anion exchange were performed as described above.

Isolation of CWP32

Heat treated H37Rv was subfractionated into subcellular fractions as described in Sørensen et al 1995. The Cell wall fraction was resuspended in 8 M urea, 0.2 % (w/v) N-octyl β-D glucopyranoside (Sigma) and 5 % (v/v) glycerol and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad) which was equilibrated with the same buffer. Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed by SDS-PAGE and fractions containing well separated bands were polled and subjected to N-terminal sequencing after transfer to PVDF membrane.

N-terminal sequencing

Fractions containing CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19B, CFP22A, CFP23A, CFP23B, CFP27, CFP30A, CWP32, and CFP50A were blotted to PVDF membrane after Tricine SDS-PAGE (Ploug et al, 1989). The relevant bands were excised and subjected to N-terminal amino acid sequence analysis on a Procise 494 sequencer (Applied Biosystems). The fraction containing CFP25A was blotted to PVDF membrane after 2-DE PAGE (isoelectric focusing in the first dimension and Tricin SDS-PAGE in the second dimension). The relevant spot was excised and sequenced as described above.

The following N-terminal sequences were obtained:

CFP7A:	AEDVRAEIVA SVLEVVVNEG DQIDKGDVVV LLESMYMEIP VLAEAAGTVS (SEQ ID NO: 81)	
CFP8A:	DPVDDAFIAKLNTAG	(SEQ ID NO: 73)
CFP8B:	DPVDAIINLDNYGX	(SEQ ID NO: 74)
CFP16:	AKLSTDELLDAFKEM	(SEQ ID NO: 79)
CFP19:	TTSPDPYAALPKLPS	(SEQ ID NO: 82)
CFP19B:	DPAXAPDVPTAAQLT	(SEQ ID NO: 80)
CFP22A:	TEYEGPKTKF HALMQ	(SEQ ID NO: 83)
CFP23A:	VIQ/AGMVT/GHIHXVAG	(SEQ ID NO: 76)
CFP23B:	AEMKXFKNAIVQEID	(SEQ ID NO: 75)
CFP25A:	AIEVSVLRVF TDSDG	(SEQ ID NO: 78)
CWP32:	TNIVVLIKQVPDTWS	(SEQ ID NO: 77)
CFP27:	TTIVALKYPG GVVMA	(SEQ ID NO: 84)
CFP30A:	SFPYFISPEX AMRE	(SEQ ID NO: 85)
CFP50:	THYDVVVLGA GPGGY	(SEQ ID NO: 86)

N-terminal homology searching in the Sanger database and identification of the corresponding genes.

The N-terminal amino acid sequence from each of the proteins was used for a homology search using the blast program of the Sanger *Mycobacterium tuberculosis* database:

http://www.sanger.ac.uk/projects/m-tuberculosis/TB-blast-server.

For CFP23B, CFP23A, and CFP19B no similarities were found in the Sanger database. This could be due to the fact that only approximately 70% of the *M. tuberculosis* genome had been sequenced when the searches were performed. The genes encoding these proteins could be contained in the remaining 30% of the genome for which no sequence data is yet available.

For CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19B , CFP22A, CFP25A, CFP27, CFP30A, CWP32, and CFP50, the following information was obtained:

CFP7A: Of the 50 determined amino acids in CFP7A a 98% identical sequence was found in cosmid csCY07D1 (contig 256):

Score = 226 (100.4 bits), Expect = 1.4e-24, P = 1.4e-24

Identities = 49/50 (98%), Positives = 49/50 (98%), Frame = -1

Query: 1 AEDVRAEIVASVLEVVVNEGDQIDKGDVVVLLESMYMEIPVLAEAAGTVS 50

AEDVRAEIVASVLEVVVNEGDQIDKGDVVVLLESM MEIPVLAEAAGTVS

Sbjct: 257679 AEDVRAEIVASVLEVVVNEGDQIDKGDVVVLLESMKMEIPVLAEAAGTVS 257530

(SEQ ID NOs: 127, 128, and 129)

The identity is found within an open reading frame of 71 amino acids length corresponding to a theoretical MW of CFP7A of 7305.9 Da and a pI of 3.762. The observed molecular weight in an SDS-PAGE gel is 7 kDa.

CFP8A: A sequence 80% identical to the 15 N-terminal amino acids was found on contig TB_1884. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 32. This gives a length of the mature protein of 98 amino acids corresponding to a theoretical MW of 9700 Da and a pI of 3.72 This is in good agreement with the observed MW on SDS-PAGE at approximately 8 kDa. The full length protein has a theoretical MW of 12989 Da and a pI of 4.38.

CFP8B: A sequence 71% identical to the 14 N-terminal amino acids was found on contig TB_653. However, careful re-evaluation of the original N-terminal sequence data confirmed the identification of the protein. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 29. This gives a length of the mature protein of 82 amino acids corresponding to a theoretical MW of 8337 Da and a pI of 4.23. This is in good agreement with the observed MW on SDS-PAGE at approximately 8 kDa. Analysis of the amino acid sequence predicts the presence of a signal peptide which has been cleaved of the mature protein found in culture filtrate.

CFP16: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY20H1.

The identity is found within an open reading frame of 130 amino acids length corresponding to a theoretical MW of CFP16 of 13440.4 Da and a pI of 4.59. The observed molecular weight in an SDS-PAGE gel is 16 kDa.

CFP19: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY270.

The identity is found within an open reading frame of 176 amino acids length corresponding to a theoretical MW of CFP19 of 18633.9 Da and a pI of 5.41. The observed molecular weight in an SDS-PAGE gel is 19 kDa.

CFP22A: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY1A6.

The identity is found within an open reading frame of 181 amino acids length corresponding to a theoretical MW of CFP22A of 20441.9 Da and a pI of 4.73. The observed molecular weight in an SDS-PAGE gel is 22 kDa.

CFP25A: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on contig 255.

The identity is found within an open reading frame of 228 amino acids length corresponding to a theoretical MW of CFP25A of 24574.3 Da and a pI of 4.95. The observed molecular weight in an SDS-PAGE gel is 25 kDa.

CFP27: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY261.

The identity is found within an open reading frame of 291 amino acids length. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 58. This gives a length of the mature protein of 233 amino acids, which corresponds to a theoretical molecular weigh at 24422.4 Da, and a theoretical pI at 4.64. The observed weight in an SDS-PAGE gel is 27 kDa.

CFP30A: Of the 13 determined amino acids in CFP30A, a 100% identical sequence was found on cosmid MTCY261.

The identity is found within an open reading frame of 248 amino acids length corresponding to a theoretical MW of CFP30A of 26881.0 Da and a pI of 5.41. The observed molecular weight in an SDS-PAGE gel is 30 kDa.

CWP32: The 15 amino acid N-terminal sequence was found to be 100% identical to a sequence found on contig 281. The identity was found within an open reading frame of 266 amino acids length, corresponding to a theoretical MW of CWP32 of 28083 Da and a pI of 4.563. The observed molecular weight in an SDS-PAGE gel is 32 kDa.

CFP50: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found in MTVO38.06. The identity is found within an open reading frame of 464 amino acids length corresponding to a theoretical MW of CFP50 of 49244 Da and a pI of 5.66. The observed molecular weight in an SDS-PAGE gel is 50 kDa.

Use of homology searching in the EMBL database for identification of CFP19A and CFP23.

Homology searching in the EMBL database (using the GCG package of the Biobase, Århus-DK) with the amino acid sequences of two earlier identified highly immunoreactive ST-CF proteins, using the TFASTA algorithm, revealed that these proteins (CFP21 and CFP25, EXAMPLE 3) belong to a family of fungal cutinase homologs. Among the most homologous sequences were also two *Mycobacterium tuberculosis* sequences found on cosmid MTCY13E12. The first, MTCY13E12.04 has 46% and 50% identity to CFP25 and CFP21 respectively. The second, MTCY13E12.05, has also 46% and 50% identity to CFP25 and CFP21. The two proteins share 62.5% aa identity in a 184 residues overlap. On the basis of the high homology to the strong T-cell antigens CFP21 and CFP25, respectively, it is believed that CFP19A and CFP23 are possible new T-cell antigens.

The first reading frame encodes a 254 amino acid protein of which the first 26 aa constitute a putative leader peptide that strongly indicates an extracellular location of the protein. The mature protein is thus 228 aa in length corresponding to a theoretical MW of 23149.0 Da and a Pi of 5.80. The protein is named CFP23.

The second reading frame encodes an 231 aa protein of which the first 44 aa constitute a putative leader peptide that strongly indicates an extracellular location of the protein. The mature protein is thus 187 aa in length corresponding to a theoretical MW of 19020.3 Da and a Pi of 7.03. The protein is named CFP19A.

The presence of putative leader peptides in both proteins (and thereby their presence in the ST-CF) is confirmed by theoretical sequence analysis using the signalP program at the Expasy molecular Biology server

(http://expasy.hcuge.ch/www/tools.html).

Searching for homologies to CFP7A, CFP16, CFP19, CFP19A, CFP19B, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32 and CFP50 in the EMBL database.

The amino acid sequences derived from the translated genes of the individual antigens were used for homology searching in the EMBL and Genbank databases using the TFASTA algorithm, in order to find homologous proteins and to address eventual functional roles of the antigens.

CFP7A: CFP7A has 44% identity and 70% similarity to hypothetical *Methanococcus jannaschii* protein (*M. jannaschii* from base 1162199-1175341), as well as 43% and 38% identity and 68 and 64% similarity to the C-terminal part of *B. stearothermophilus* pyruvate carboxylase *and Streptococcus mutans* biotin carboxyl carrier protein.

CFP7A contains a consensus sequence EAMKM for a biotin binding site motif which in this case was slightly modified (ESMKM in amino acid residues 34 to 38). By incubation with alkaline phosphatase conjugated streptavidin after SDS-PAGE and transfer to nitrocellulose it was demonstrated that native CFP7A was biotinylated.

CFP16: RplL gene, 130 aa. Identical to the *M. bovis* 50s ribosomal protein L7/L12 (acc. No P37381).

CFP19: CFP19 has 47% identity and 55% similarity to *E.coli* pectinesterase homolog (ybhC gene) in a 150 aa overlap.

CFP19A: CFP19A has between 38% and 45% identity to several cutinases from different fungal sp.

In addition CFP19A has 46% identity and 61% similarity to CFP25 as well as 50% identity and 64% similarity to CFP21 (both proteins are earlier isolated from the ST-CF).

CFP19B: No apparent homology

CFP22A: No apparent homology

CFP23: CFP23 has between 38% and 46% identity to several cutinases from different fungal sp.

In addition CFP23 has 46% identity and 61% similarity to CFP25 as well as 50% identity and 63% similarity to CFP21 (both proteins are earlier isolated from the ST-CF).

CFP25A: CFP25A has 95% identity in a 241 aa overlap to a putative *M. tuberculosis* thymidylate synthase (450 aa accession No p28176).

CFP27: CFP27 has 81% identity to a hypothetical *M. leprae* protein and 64% identity and 78% similarity to *Rhodococcus* sp. proteasome beta-type subunit 2 (prcB(2) gene).

CFP30A: CFP30A has 67% identity to *Rhodococcus proteasome* alfa-type 1 subunit.

CWP32: The CWP32 N-terminal sequence is 100% identical to the *Mycobacterium leprae* sequence MLCB637.03.

CFP50: The CFP50 N-terminal sequence is 100% identical to a putative lipoamide dehydrogenase from *M. leprae* (Accession 415183)

Cloning of the genes encoding CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19A, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32, and CFP50.

The genes encoding CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19A, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32 and CFP50 were all cloned into the expression vector pMCT6, by PCR amplification with gene specific primers, for recombinant expression in *E. coli* of the proteins.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1X low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 ml reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluescript SK II+ - T vector (Stratagene). Plasmid DNA was hereafter prepared from clones harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMCT6 in frame with 8 histidines which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

For cloning of the individual antigens, the following gene specific primers were used:

CFP7A: Primers used for cloning of *cfp7A*:

OPBR-79: AAGAGTAGATCTATGATGGCCGAGGATGTTCGCG (SEQ ID NO: 95)
OPBR-80: CGGCGACGACGGATCCTACCGCGTCGG (SEQ ID NO: 96)

OPBR-79 and OPBR-80 create *Bgl*II and *Bam*HI sites, respectively, used for the cloning in pMCT6.

CFP8A: Primers used for cloning of *cfp*8A:

CFP8A-F: CTGAGATCTATGAACCTACGGCGCC (SEQ ID NO: 154)
CFP8A-R: CTCCCATGGTACCCTAGGACCCGGGCAGCCCCGGC (SEQ ID NO: 155)

CFP8A-F and CFP8A-R create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP8B: Primers used for cloning of *cfp*8B:

CFP8B-F: CTGAGATCTATGAGGCTGTCGTTGACCGC (SEQ ID NO: 156)
CFP8B-R: CTCCCCGGGCTTAATAGTTGTTGCAGGAGC (SEQ ID NO: 157)

CFP8B-F and CFP8B-R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP16: Primers used for cloning of *cfp*16:

OPBR-104: CCGGGAGATCTATGGCAAAGCTCTCCACCGACG (SEQ ID NOs: 111 and 130)
OPBR-105: CGCTGGGCAGAGCTACTTGACGGTGACGGTGG (SEQ ID NOs: 112 and 131)

OPBR-104 and OPBR-105 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP19: Primers used for cloning of *cfp*19:

OPBR-96: GAGGAAGATCTATGACAACTTCACCCGACCCG (SEQ ID NO: 107)
OPBR-97. CATGAAGCCATGGCCCGCAGGCTGCATG (SEQ ID NO: 108)

OPBR-96 and OPBR-97 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP19A: Primers used for cloning of *cfp*19A:

OPBR-88: CCCCCCAGATCTGCACCACCGGCATCGGCGGGC (SEQ ID NO: 99)
OPBR-89. GCGGCGGATCCGTTGCTTAGCCGG (SEQ ID NO: 100)

OPBR-88 and OPBR-89 create *Bgl*II and *Bam*HI sites, respectively, used for the cloning in pMCT6.

CFP22A: Primers used for cloning of *cfp*22A:

OPBR-90: CCGGCTGAGATCTATGACAGAATACGAAGGGC (SEQ ID NO: 101)
OPBR-91: CCCCGCCAGGGAACTAGAGGCGGC (SEQ ID NO: 102)

OPBR-90 and OPBR-91 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP23: Primers used for cloning of *cfp*23:

OPBR-86: CCTTGGGAGATCTTTGGACCCCGGTTGC (SEQ ID NO: 97)
OPBR-87: GACGAGATCTTATGGGCTTACTGAC (SEQ ID NO: 98)

OPBR-86 and OPBR-87 both create a *Bgl*II site used for the cloning in pMCT6.

CFP25A: Primers used for cloning of *cfp*25A:

OPBR-106: GGCCCAGATCTATGGCCATTGAGGTTTCGGTGTTGC (SEQ ID NO: 113)
OPBR-107: CGCCGTGTTGCATGGCAGCGCTGAGC (SEQ ID NO: 114)

OPBR-106 and OPBR-107 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP27: Primers used for cloning of *cfp*27:

OPBR-92: CTGCCGAGATCTACCACCATTGTCGCGCTGAAATACCC (SEQ ID NO: 103)

OPBR-93: CGCCATGGCCTTACGCGCCAACTCG (SEQ ID NO: 104)

OPBR-92 and OPBR-93 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP30A: Primers used for cloning of *cfp*30A:

OPBR-94: GGCGGAGATCTGTGAGTTTTCCGTATTTCATC (SEQ ID NO: 105)
OPBR-95: CGCGTCGAGCCATGGTTAGGCGCAG (SEQ ID NO: 106)

OPBR-94 and OPBR-95 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CWP32: Primers used for cloning of *cwp*32:

CWP32-F: GCTTAGATCTATGATTTTCTGGGCAACCAGGTA (SEQ ID NO: 158)
CWP32-R: GCTTCCATGGGCGAGGCACAGGCGTGGGAA (SEQ ID NO: 159)

CWP32-F and CWP32-R create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP50: Primers used for cloning of *cfp*50:

OPBR-100: GGCCGAGATCTGTGACCCACTATGACGTCGTCG (SEQ ID NO: 109)
OPBR-101: GGCGCCCATGGTCAGAAATTGATCATGTGGCCAA (SEQ ID NO: 110)

OPBR-100 and OPBR-101 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

Expression/purification of recombinant CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19A, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32, and CFP50 proteins.

Expression and metal affinity purification of recombinant proteins was undertaken essentially as described by the manufacturers. For each protein, 1 l LB-media containing 100 µg/ml ampicillin,

was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmids. Cultures were shaken at 37°C until they reached a density of OD_{600} = 0.4 - 0.6. IPTG was hereafter added to a final concentration of 1 mM and the cultures were further incubated 4-16 hours. Cells were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.

After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analyzed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content were determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 6

Identification of CFP7B, CFP10A, CFP11 and CFP30B.

Isolation of CFP7B

ST-CF was precipitated with ammonium sulphate at 80% saturation and redissolved in PBS, pH 7.4, and dialyzed 3 times against 25 mM Piperazin-HCl, pH 5.5, and subjected to croma-tofocusing on a matrix of PBE 94 (Pharmacia) in a column connected to an FPLC system (Pharmacia). The column was equilibrated with 25 mM Piperazin-HCl, pH 5.5, and the elution was performed with 10% PB74-HCl, pH 4.0 (Pharmacia). Fractions with similar band patterns

were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml. An equal volume of SDS containing sample buffer was added and the protein solution boiled for 5 min before further separation on a Multi-Eluter (BioRad) in a matrix of 10-20 % polyacrylamid (Andersen,P. & Heron,I., 1993). The fraction containing a well separated band below 10 kDa was selected for N-terminal sequencing after transfer to a PVDF membrane.

Isolation of CFP11

ST-CF was precipitated with ammonium sulphate at 80% saturation. The precipitated proteins were removed by centrifugation and after resuspension washed with 8 M urea. CHAPS and glycerol were added to a final concentration of 0.5 % (w/v) and 5% (v/v) respectively and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad). The Rotofor Cell had been equilibrated with an 8M urea buffer containing 0.5 % (w/v) CHAPS, 5% (v/v) glycerol, 3% (v/v) Biolyt 3/5 and 1% (v/v) Biolyt 4/6 (BioRad). Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed on silver-stained 10-20% SDS-PAGE. The fractions in the pH gradient 5.5 to 6 were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1 ml. 300 mg of the protein preparation was separated on a 10-20% Tricine SDS-PAGE (Ploug et al 1989) and transferred to a PVDF membrane and Coomassie stained. The lowest band occurring on the membrane was excised and submitted for N-terminal sequencing.

Isolation of CFP10A and CFP30B

ST-CF was concentrated approximately 10-fold by ultrafiltration and ammonium sulphate precipitation at 80 % saturation. Proteins were redissolved in 50 mM sodium phosphate, 1.5 M ammonium sulphate, pH 8.5, and subjected to thiophilic adsorption chromatography on an Affi-T gel column (Kem-En-Tec). Proteins were eluted by a 1.5 to 0 M decreasing gradient of ammonium sulphate. Fractions with similar band patterns in SDS-PAGE were pooled and anion exchange chromatography was performed on a Mono Q HR 5/5 column connected to an FPLC system (Pharmacia). The column was equilibrated with 10 mM Tris-HCl, pH 8.5, and the elution was performed with a gradient of NaCl from 0 to 1 M. Fractions containing well separated bands in SDS-PAGE were selected.

Fractions containing CFP10A and CFP30B were blotted to PVDF membrane after 2-DE PAGE (Ploug et al, 1989). The relevant spots were excised and subjected to N-terminal amino acid sequence analysis.

N-terminal sequencing

N-terminal amino acid sequence analysis was performed on a Procise 494 sequencer (applied Biosystems).

The following N-terminal sequences were obtained:

CFP7B:	PQGTVKWFNAEKGFG	(SEQ ID NO: 168)
CFP10A:	NVTVSIPTILRPXXX	(SEQ ID NO: 169)
CFP11:	TRFMTDPHAMRDMAG	(SEQ ID NO: 170)
CFP30B:	PKRSEYRQGTPNWVD	(SEQ ID NO: 171)

"X" denotes an amino acid which could not be determined by the sequencing method used.

N-terminal homology searching in the Sanger database and identification of the corresponding genes.

The N-terminal amino acid sequence from each of the proteins was used for a homology search using the blast program of the Sanger *Mycobacterium tuberculosis* genome database:

http//www.sanger.ac.uk/projects/m-tuberculosis/TB-blast-server.

For CFP11 a sequence 100% identical to 15 N-terminal amino acids was found on contig TB_1314. The identity was found within an open reading frame of 98 amino acids length corresponding to a theoretical MW of 10977 Da and a pI of 5.14.

Amino acid number one can also be an Ala (insted of a Thr) as this sequence was also obtained (results not shown), and a 100% identical sequence to this N-terminal is found on contig TB_671 and on locus MTCI364.09.

For CFP7B a sequence 100% identical to 15 N-terminal amino acids was found on contig TB_2044 and on locus MTY15C10.04 with EMBL accession number: z95436. The identity was found within an open reading frame of 67 amino acids length corresponding to a theoretical MW of 7240 Da and a pI of 5.18.

For CFP10A a sequence 100% identical to 12 N-terminal amino acids was found on contig TB_752 and on locus CY130.20 with EMBL accession number: Q10646 and Z73902. The identity was found within an open reading frame of 93 amino acids length corresponding to a theoretical MW of 9557 Da and a pI of 4.78.

For CFP30B a sequence 100% identical to 15 N-terminal amino acids was found on contig TB_335. The identity was found within an open reading frame of 261 amino acids length corresponding to a theoretical MW of 27345 Da and a pI of 4.24.

The amino acid sequences of the purified antigens as picked from the Sanger database are shown in the following list.

CFP7B (SEQ ID NO: 147)

1 MPQGTVKWFN AEKGFGFIAP EDGSADVFVH YTEIQGTGFR TLEENQKVEF
51 EIGHSPKGPQ ATGVRSL

CFP10A (SEQ ID NO: 141)

1 MNVTVSIPTI LRPHTGGQKS VSASGDTLGA VISDLEANYS GISERLMDPS
51 SPGKLHRFVN IYVNDEDVRF SGGLATAIAD GDSVTILPAV AGG

CFP11 protein sequence (SEQ ID NO: 143)

1 MATRFMTDPH AMRDMAGRFE VHAQTVEDEA RRMWASAQNI SGAGWSGMAE
51 ATSLDTMAQM NQAFRNIVNM LHGVRDGLVR DANNYEQQEQ ASQQILSS

CFP30B (SEQ ID NO: 145)

1 MPKRSEYRQG TPNWVDLQTT DQSAAKKFYT SLFGWGYDDN PVPGGGGVYS
51 MATLNGEAVA AIAPMPPGAP EGMPPIWNTY IAVDDVDAVV DKVVPGGGQV
101 MMPAFDIGDA GRMSFITDPT GAAVGLWQAN RHIGATLVNE TGTLIWNELL
151 TDKPDLALAF YEAVVGLTHS SMEIAAGQNY RVLKAGDAEV GGCMEPPMPG
201 VPNHWHVYFA VDDADATAAK AAAAGGQVIA EPADIPSVGR FAVLSDPQGA
251 IFSVLKPAPQ Q

Cloning of the genes encoding CFP7B, CFP10A, CFP11, and CFP30B.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1X low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 ml reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec., using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluscript SK II+ - T vector (Stratagene). Plasmid DNA was hereafter prepared from clones harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMCT6 in frame with 8 histidines which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

For cloning of the individual antigens, the following gene specific primers were used:

CFP7B: Primers used for cloning of *cfp*7B:

CFP7B-F: CTGAGATCTAGAATGCCACAGGGAACTGTG (SEQ ID NO: 160)
CFP7B-R: TCTCCCGGGGGTAACTCAGAGCGAGCGGAC (SEQ ID NO: 161)

CFP7B-F and CFP7B-R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP10A: Primers used for cloning of *cfp*10A:

CFP10A-F: CTGAGATCTATGAACGTCACCGTATCC (SEQ ID NO: 162)
CFP10A-R: TCTCCCGGGGCTCACCCACCGGCCACG (SEQ ID NO: 163)

CFP10A -F and CFP10A -R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP11: Primers used for cloning of *cfp*11:

CFP11-F: CTGAGATCTATGGCAACACGTTTTATGACG (SEQ ID NO: 164)
CFP11-R: CTCCCCGGGTTAGCTGCTGAGGATCTGCTH (SEQ ID NO: 165)

CFP11-F and CFP11-R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP30B: Primers used for cloning of *cfp*30B:

CFP30B-F: CTGAAGATCTATGCCCAAGAGAAGCGAATAC (SEQ ID NO: 166)
CFP30B -R: CGGCAGCTGCTAGCATTCTCCGAATCTGCCG (SEQ ID NO: 167)

CFP30B-F and CFP30B-R create *Bgl*II and *Pvu*II sites, respectively, used for the cloning in pMCT6.

Expression/purification of recombinant CFP7B, CFP10A, CFP11 and CFP30B protein.

Expression and metal affinity purification of recombinant protein was undertaken essentially as described by the manufacturers. 1 l LB-media containing 100 µg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmid. The culture was shaken at 37 °C until it reached a density of OD_{600} = 0.5. IPTG was hereafter added to a final concentration of 1 mM and the culture was further incubated 4 hours. Cells

were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.

After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analysed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content was determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 7

Using homology searching for identification of ORF11-1, ORF11-2, ORF11-3 and ORF11-4.

A search of the *Mycobacterium tuberculosis* Sanger sequence database with the amino acid sequences of CFP11, a previously identified ST-CF protein, identified 4 new very homologous proteins. All 4 proteins were at least 96% homologous to CFP11.

On the basis of the strong homology to CFP11, it is belived that ORF11-1, ORF11-2, ORF11-3 and ORF11-4 are potential new T-cell antigens.

The first open reading frame, MTCY10G2.11, homologous to CFP11, encodes a protein of 98 amino acids corresponding to a theoretical molecular mass of 10994Da and a pI of 5.14. The protein was named ORF11-1.

The second open reading frame, MTCI364.09, homologous to CFP11, encodes a protein of 98 amino acids corresponding to a theoretical molecular mass of 10964Da and a pI of 5.14. The protein was named ORF11-2.

The third open reading frame, MTV049.14, has an in frame stop codon. Because of the very conserved DNA sequence in this position amongst the 4 open reading frames it is however suggested that this is due to a sequence mistake.
The "T" in position 175 of the DNA sequence is therefor suggested to be a "C" as in the four other ORF's. The Q in position 59 in the amino acid sequence would have been a "stop" if the T in position 175 in the DNA sequence had not been substituted.
The open reading frame encodes a protein of 98 amino acids corresponding to a theoretical molecular mass of 10994Da and a pI of 5.14. The protein was named ORF11-3.

The fourth open reading frame, MTCY15C10.32, homologous to CFP11, encodes a protein of 98 amino acids corresponding to a theoretical molecular mass of 11024Da and a pI of 5.14. The protein was named ORF11-4.

Using homology searching for identification of ORF7-1 and ORF7-2.

A search of the *Mycobacterium tuberculosis* Sanger sequence database with the amino acid sequences of a previously identified immunoreactive ST-CF protein, CFP7, identified 2 new very homologous proteins. The protein ORF7-1 (MTV012.33) was 84% identical to CFP7, with a primary structure of the same size as CFP7, and the protein ORF7-2 (MTV012.31) was 68% identical to CFP7 in a 69 amino acid overlap.
On the basis of the strong homology to the potent human T-cell antigen CFP7, ORF7-1 and ORF7-2 are belived to be potential new T-cell antigens.

The first open reading frame homologous to CFP7, encodes a protein of 96 amino acids corresponding to a theoretical molecular mass of 10313Da and a pI of 4.186. The protein was named ORF7-1.

The second open reading frame homologous to CFP7, encodes a protein of 120 amino acids corresponding to a theoretical molecular mass of 12923.00 Da and a pI of 7.889. The protein was named ORF7-2.

Cloning of the homologous *orf7-1* and *orf7-2.*

Since *ORF7-1* and *ORF7-2* are nearly identical to CFP7 it was nessesary to use the flanking DNA regions in the cloning procedure, to ensure the cloning of the correct ORF. Two PCR reactions were carried out with two different primer sets. PCR reaction 1 was carried out using *M. tuberculosis* chromosomal DNA and a primerset corresponding to the flanking DNA. PCR reaction 2 was carried out directly on the first PCR product using ORF specific primers which introduced restriction sites for use in the later cloning procedure.
The sequences of the primers used are given below;

Orf7-1:

Primers used for the initial PCR reaction (1) using *M. tuberculosis* chromosomal DNA as template;

Sence: MTV012.33-R1: 5'- GGAATGAAAAGGGGTTTGTG - 3' (SEQ ID NO: 186)
Antisence:MTV012.33-F1: 5'- GACCACGCCCGCGCCGTGTG - 3'(SEQ ID NO:187)

Primers used for the second round of PCR (2) using PCR product 1 as template;

Sence: MTV012.33-R2: 5' - ***GCAACACCCGGG***ATGTCGCAGATTATG - 3'
(SEQ ID NO: 188)
(introduces a *Sma*I upstream of the *orf7-1* start codon)
Antisence:MTV012.33-F2: 5' - ***CTAAGCTTGGATCC***CTAGCCGCCCCACTTG - 3' ((SEQ ID NO: 189)
(introduces a *Bam*HI downstream of the *orf7-1* stop codon).

Orf7-2:

Primers used for the initial PCR reaction (1) using *M. tuberculosis* chromosomal DNA as template;

Sence: MTV012.31-R1: 5'- GAATATTTGAAAGGGATTCGTG - 3' (SEQ ID NO: 190)

Antisence:MTV012.31-F1: 5'- ***CTACTAAGCTTGGATCC***TTAGTCTCCGGCG - 3'
(SEQ ID NO: 191)
(introduces a *Bam*HI downstream of the *orf7-2* stop codon)

Primers used for the second round of PCR (2) using PCR product 1 as template;

Sence: MTV012.31-R2: 5' -***GCAACACCCGGG***GTGTCGCAGAGTATG- 3'
(SEQ ID NO: 192)
(introduces a *Sma*I upstream of the *orf7-2* start codon)
Antisence:MTV012.31-F1: 5'- ***CTACTAAGCTTGGATCC***TTAGTCTCCGGCG - 3'
(SEQ ID NO: 193)
(introduces a BamHI downstream of the orf7-2 stop codon)

The genes encoding ORF7-1 and ORF7-2 were cloned into the expression vector pMST24, by PCR amplification with gene specific primers, for recombinant expression in *E. coli* of the proteins.
The first PCR reactions contained either 10 ng of *M. tuberculosis* chromosomal DNA (PCR reaction 1) or 10ng PCR product 1 (PCR reaction 2) in 1 x low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Tag+ DNA polymerase (Stratagene) in 10 ml reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.
The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluscript SK II+ - T vector (Stratagene). Plasmid DNA was hereafter prepared from clones harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMST24 in frame with 6 histidines which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

Expression/purification of recombinant ORF7-1 and ORF7-2 protein.

Expression and metal affinity purification of recombinant protein was undertaken essentially as described by the manufacturers. 1 l LB-media containing 100 µg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmid. The culture was shaken at 37 °C until it reached a density of OD600 = 0.5. IPTG was hereafter added to a final concentration of 1 mM and the culture was further incubated 2 hours. Cells were harvested, resuspended in 1 X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.

After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analysed by SDS-PAGE. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content was determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 8

Cloning of the gene expressing CFP26 (MPT51)

Synthesis and design of probes

Oligonucleotide primers were synthesized automatically on a DNA synthesizer (Applied Biosystems, Forster City, Ca, ABI-391, PCR-mode) deblocked and purified by ethanol precipitation.

Three oligonucleotides were synthesized (TABLE 3) on the basis of the nucleotide sequence from *mpb51* described by Ohara *et al.* (1995). The oligonucleotides were engineered to include

an *Eco*RI restriction enzyme site at the 5' end and at the 3' end by which a later subcloning was possible.

Additional four oligonucleotides were synthesized on the basis of the nucleotide sequence from MPT51 (Fig. 5 and SEQ ID NO: 41). The four combinations of the primers were used for the PCR studies.

DNA cloning and PCR technology

Standard procedures were used for the preparation and handling of DNA (Sambrook *et al.*, 1989). The gene *mpt51* was cloned from *M. tuberculosis* H37Rv chromosomal DNA by the use of the polymerase chain reactions (PCR) technology as described previously (Oettinger and Andersen, 1994). The PCR product was cloned in the pBluescriptSK + (Stratagene).

Cloning of *mpt51*

The gene, the signal sequence and the Shine Delgarno region of MPT51 was cloned by use of the PCR technology as two fragments of 952 bp and 815 bp in pBluescript SK +, designated pTO52 and pTO53.

DNA Sequencing

The nucleotide sequence of the cloned 952 bp *M. tuberculosis* H37Rv PCR fragment, pTO52, containing the Shine Dalgarno sequence, the signal peptide sequence and the structural gene of MPT51, and the nucleotide sequence of the cloned 815 bp PCR fragment containing the structural gene of MPT51, pTO53, were determined by the dideoxy chain termination method adapted for supercoiled DNA by use of the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., Cleveland, OH) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

The nucleotide sequences of pTO52 and pTO53 and the deduced amino acid sequence are shown in Figure 5. The DNA sequence contained an open reading frame starting with a ATG codon at position 45 - 47 and ending with a termination codon (TAA) at position 942 - 944. The

nucleotide sequence of the first 33 codons was expected to encode the signal sequence. On the basis of the known N-terminal amino acid sequence (Ala - Pro - Tyr - Glu - Asn) of the purified MPT51 (Nagai *et al.*, 1991) and the features of the signal peptide, it is presumed that the signal peptidase recognition sequence (Ala-X-Ala) (von Heijne, 1984) is located in front of the N-terminal region of the mature protein at position 144. Therefore, a structural gene encoding MPT51, *mpt51*, derived from *M. tuberculosis* H37Rv was found to be located at position 144 - 945 of the sequence shown in Fig. 11. The nucleotide sequence of *mpt51* differed with one nucleotide compared to the nucleotide sequence of MPB51 described by Ohara *et al.* (1995) (Fig. 11). In *mpt51* at position 780 was found a substitution of a guanine to an adenine. From the deduced amino acid sequence this change occurs at a first position of the codon giving a amino acid change from alanine to threonine. Thus it is concluded, that *mpt51* consists of 801 bp and that the deduced amino acid sequence contains 266 residues with a molecular weight of 27,842, and MPT51 show 99,8% identity to MPB51.

Subcloning of *mpt51*

An *Eco*RI site was engineered immediately 5' of the first codon of *mpt51* so that only the coding region of the gene encoding MPT51 would be expressed, and an *Eco*RI site was incorporated right after the stop codon at the 3' end.

DNA of the recombinant plasmid pTO53 was cleaved at the *Eco*RI sites. The 815 bp fragment was purified from an agarose gel and subcloned into the *Eco*RI site of the pMAL-*cR1* expression vector (New England Biolabs), pTO54. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue by the standard procedures for DNA manipulation.

The endpoints of the gene fusion were determined by the dideoxy chain termination method as described under section DNA sequencing. Both strands of the DNA were sequenced.

Preparation and purification of rMPT51

Recombinant antigen was prepared in accordance with instructions provided by New England Biolabs. Briefly, single colonies of *E. coli* harbouring the pTO54 plasmid were inoculated into Luria-Bertani broth containing 50 µg/ml ampicillin and 12.5 µg/ml tetracycline and grown at 37°C to 2 x 10^8 cells/ml. Isopropyl-β-D-thiogalactoside (IPTG) was then added to a final concentration

of 0.3 mM and growth was continued for further 2 hours. The pelleted bacteria were stored overnight at -20°C in new column buffer (20 mM Tris/HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT))and thawed at 4°C followed by incubation with 1 mg/ml lysozyme on ice for 30 min and sonication (20 times for 10 sec with intervals of 20 sec). After centrifugation at 9,000 x g for 30 min at 4°C, the maltose binding protein -MPT51fusion protein (MBP-rMPT51) was purified from the crude extract by affinity chromatography on amylose resin column. MBP-rMPT51 binds to amylose. After extensive washes of the column, the fusion protein was eluted with 10 mM maltose. Aliquots of the fractions were analyzed on 10% SDS-PAGE. Fractions containing the fusion protein of interest were pooled and was dialysed extensively against physiological saline.

Protein concentration was determined by the BCA method supplied by Pierce (Pierce Chemical Company, Rockford, IL).

TABLE 3. Sequence of the *mpt51* oligonucleotides[a].

Orientation and oligonucleotide[a]	Sequences (5'→ 3')	Position[b] (nucleotide)
Sense		
MPT51-1	CTCGAATTCGCCGGGTGCACACAG (SEQ ID NO: 28)	6 - 21 (SEQ ID NO: 41)
MPT51-3	CTCGAATTCGCCCCATACGAGAAC (SEQ ID NO: 29)	143 - 158 (SEQ ID NO: 41)
MPT51-5	GTGTATCTGCTGGAC (SEQ ID NO: 30)	228 - 242 (SEQ ID NO: 41)
MPT51-7	CCGACTGGCTGGCCG (SEQ ID NO: 31)	418 - 432 (SEQ ID NO: 41)
Antisense		
MPT51-2	GAGGAATTCGCTTAGCGGATCGCA (SEQ ID NO: 32)	946 - 932 (SEQ ID NO: 41)
MPT51-4	CCCACATTCCGTTGG (SEQ ID NO: 33)	642 - 628 (SEQ ID NO: 41)
MPT51-6	GTCCAGCAGATACAC (SEQ ID NO: 34)	242 - 228 (SEQ ID NO: 41)

[a] The oligonucleotides MPT51-1 and MPT51-2 were constructed from the MPB51 nucleotide sequence (Ohara *et al.*, 1995). The other oligonucleotides constructions were based on the nucleotide sequence obtained from *mpt51* reported in this work. Nucleotides (nt) underlined are not contained in the nucleotide sequence of MPB/T51.

[b] The positions referred to are of the non-underlined parts of the primers and correspond to the nucleotide sequence shown in SEQ ID NO: 41.

Cloning of *mpt51* in the expression vector pMST24.

A PCR fragment was produced from pTO52 using the primer combination MPT51-F and MPT51-R (TABLE 4). A *Bam*HI site was engineered immediately 5' of the first codon of *mpt51* so that only the coding region of the gene encoding MPT51 would be expressed, and an *Nco*I site was incorporated right after the stop codon at the 3' end.

The PCR product was cleaved at the *Bam*HI and the *Nco*I site. The 811 bp fragment was purified from an agarose gel and subcloned into the *Bam*HI and the *Nco*I site of the pMST24 expression vector, pTO86. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue by the standard procedures for DNA manipulation.

The nucleotide sequence of complete gene fusion was determined by the dideoxy chain termination method as described under section DNA sequencing. Both strands of the DNA were sequenced.

Preparation and purification of rMPT51.

Recombinant antigen was prepared from single colonies of *E. coli* harbouring the pTO86 plasmid inoculated into Luria-Bertani broth containing 50 µg/ml ampicillin and 12.5 µg/ml tetracycline and grown at 37°C to 2 x 10^8 cells/ml. Isopropyl-β-D-thiogalactoside (IPTG) was then added to a final concentration of 1 mM and growth was continued for further 2 hours. The pelleted bacteria were resuspended in BC 100/20 buffer (100 mM KCl, 20 mM Imidazole, 20 mM Tris/HCl, pH 7.9, 20 % glycerol). Cells were broken by sonication (20 times for 10 sec with intervals of 20 sec). After centrifugation at 9,000 x g for 30 min. at 4°C the insoluble matter was resuspended in BC 100/20 buffer with 8 M urea followed by sonication and centrifugation as above. The 6 x His tag-MPT51 fusion protein (His-rMPT51) was purified by affinity chromatography on Ni-NTA resin column (Qiagen, Hilden, Germany). His-rMPT51 binds to Ni-NTA. After extensive washes of the column, the fusion protein was eluted with BC 100/40 buffer (100 mM KCl, 40 mM Imidazole, 20 mM Tris/HCl, pH 7.9, 20 % glycerol) with 8 M urea and BC 1000/40 buffer (1000 mM KCl, 40 mM Imidazole, 20 mM Tris/HCl, pH 7.9, 20 % glycerol) with 8 M urea. His-rMPT51 was extensive dialysed against 10 mM Tris/HCl, pH 8.5, 3 M urea followed by purification using fast protein liquid chromatography (FPLC) (Pharmacia, Uppsala, Sweden), over an anion exchange column (Mono Q) using 10 mM Tris/HCl, pH 8.5, 3 M urea with a 0 - 1 M NaCl linear gradient. Fractions containing rMPT51 were pooled and subsequently dialysed extensively against 25 mM Hepes, pH 8.0 before use.

Protein concentration was determined by the BCA method supplied by Pierce (Pierce Chemical Company, Rockford, IL).

The lipopolysaccharide (LPS) content was determined by the limulus amoebocyte lysate test (LAL) to be less than 0.004 ng/µg rMPT51, and this concentration had no influence on cellular activity.

TABLE 4. Sequence of the *mpt51* oligonucleotides.

Orientation and oligonucleotide	Sequences (5' → 3')	Position (nt)
Sense		
MPT51-F	CTCGGATCCTGCCCCATACGAGAACCTG	139 - 156
Antisense		
MPT51-R	CTCCCATGGTTAGCGGATCGCACCG	939 - 924

EXAMPLE 9:

Mapping of the purified antigens in a 2DE system.

In order to characterize the purified antigens they were mapped in a 2-dimensional electrophoresis (2DE) reference system. This consists of a silver stained gel containing ST-CF proteins separated by isoelectrical focusing followed by a separation according to size in a polyacrylamide gel electrophoresis. The 2DE was performed according to Hochstrasser *et al.* (1988). 85 µg of ST-CF was applied to the isoelectrical focusing tubes where BioRad ampholytes BioLyt 4-6 (2 parts) and BioLyt 5-7 (3 parts) were included. The first dimension was performed in acrylamide/piperazin diacrylamide tube gels in the presence of urea, the detergent CHAPS and the reducing agent DTT at 400 V for 18 hours and 800 V for 2 hours. The second dimension 10-20% SDS-PAGE was performed at 100 V for 18 hours and silver stained. The identification of CFP7, CFP7A, CFP7B, CFP8A, CFP8B, CFP9, CFP11, CFP16, CFP17, CFP19, CFP20, CFP21, CFP22, CFP25, CFP27, CFP28, CFP29, CFP30A, CFP50, and MPT51 in the 2DE reference gel were done by comparing the spot pattern of the purified antigen with ST-CF with and without the purified antigen. By the assistance of an analytical 2DE software system (Phoretix International, UK) the spots have been identified in Fig. 12. The position of MPT51 and CFP29 were confirmed by a Western blot of the 2DE gel using the Mab's anti-CFP29 and HBT 4.

EXAMPLE 10

Biological activity of the purified antigens.

IFN-γ induction in the mouse model of TB infection

The recognition of the purified antigens in the mouse model of memory immunity to TB (described in example 1) was investigated. The results shown in TABLE 5 are representative for three experiments.

A very high IFN-γ response was induced by two of the antigens CFP17 and CFP21 at almost the same high level as ST-CF.

TABLE 5. IFN-γ release from splenic memory effector cells from C57BL/6J mice isolated after reinfection with *M. tuberculosis* after stimulation with native antigens.

Antigen[a]	IFN-γ (pg/ml)[b]
ST-CF	12564
CFP7	ND[d]
CFP9	ND
CFP17	9251
CFP20	2388
CFP21	10732
CFP22 + CFP25[c]	5342
CFP26 (MPT51)	ND
CFP28	2818
CFP29	3700

The data is derived from a representative experiment out of three.

[a] ST-CF was tested in a concentration of 5 µg/ml and the individual antigens in a concentration of 2 µg/ml.

[b] Four days after rechallenge a pool of cells from three mice were tested. The results are expressed as mean of duplicate values and the difference between duplicate cultures are < 15% of mean. The IFN-γ release of cultures incubated without antigen was 390 pg/ml.

[c] A pool of CFP22 and CFP25 was tested.

[d] ND, not determined.

Skin test reaction in TB infected guinea pigs

The skin test activity of the purified proteins was tested in *M. tuberculosis* infected guinea pigs.

1 group of guinea pigs was infected via an ear vein with 1 x 10^4 CFU of *M. tuberculosis* H37Rv in 0,2 ml PBS. After 4 weeks skin tests were performed and 24 hours after injection erythema diameter was measured.

As seen in TABLES 6 and 6a all of the antigens induced a significant Delayed Type Hypersensitivity (DTH) reaction.

TABLE 6. DTH erythema diameter in guinea pigs infected with 1 x 10^4 CFU of *M. tuberculosis*, after stimulation with native antigens.

Antigen[a]	Skin reaction (mm)[b]
Control	2.00
PPD[c]	15.40 (0.53)
CFP7	ND[e]
CFP9	ND
CFP17	11.25 (0.84)
CFP20	8.88 (0.13)
CFP21	12.44 (0.79)
CFP22 + CFP25[d]	9.19 (3.10)
CFP26 (MPT51)	ND
CFP28	2.90 (1.28)
CFP29	6.63 (0.88)

The values presented are the mean of erythema diameter of four animals and the SEM's are indicated in the brackets. For PPD and CFP29 the values are mean of erythema diameter of ten animals.

[a] The antigens were tested in a concentration of 0,1 µg except for CFP29 which was tested in a concentration of 0,8 µg.

[b] The skin reactions are measured in mm erythema 24 h after intradermal injection.

[c] 10 TU of PPD was used.

[d] A pool of CFP22 and CFP25 was tested.

[e] ND, not determined.

Together these analyses indicate that most of the antigens identified were highly biologically active and recognized during TB infection in different animal models.

TABLE 6a. DTH erythema diameter of recombinant antigens in outbred guinea pigs infected with 1×10^4 CFU of *M. Tuberculosis*.

Antigen[a]	Skin reaction (mm)[b]
Control	2.9 (0.3)
PPD[c]	14.5 (1.0)
CFP 7a	13.6 (1.4)
CFP 17	6.8 (1.9)
CFP 20	6.4 (1.4)
CFP 21	5.3 (0.7)
CFP 25	10.8 (0.8)
CFP 29	7.4 (2.2)
MPT 51	4.9 (1.1)

The values presented are the mean of erythema diameter of four animals and the SEM's are indicated in the brackets. For Control, PPD, and CFP 20 the values are mean of erythema diameter of eight animals.

[a] The antigens were tested in a concentration of 1,0 µg.

[b] The skin test reactions are measured in mm erythema 24 h after intradermal infection.

[c] 10 TU of PPD was used.

Table 6B.

DTH erythema diameter in guinea pigs i.v. infected with 1×10^4 CFU *M. tuberculosis*, after stimulation with 10 µg antigen.

Antigen	Mean (mm)	SEM
PBS	3,25	0,48
PPD (2TU)	10,88	1
nCFP7B	7,0	0,46
nCFP19	6,5	0,74

The values presented are the mean of erythema diameter of four animals.

The results in Table 6B indicates biological activity of nCFP7B and nCFP19.

Biological activity of the purified recombinant antigens.

Interferon-γ induction in the mouse model of TB infection.

Primary infections. 8 to 12 weeks old female C57BL/6j(H-2^b), CBA/J(H-2^k), DBA.2(H-2^d) and A.SW(H-2^s) mice (Bomholtegaard, Ry) were given intravenous infections via the lateral tail vein with an inoculum of 5 x 10^4 *M. tuberculosis* suspended in PBS in a vol. of 0.1 ml. 14 days postinfection the animals were sacrificed and spleen cells were isolated and tested for the recognition of recombinant antigen.

As seen in TABLE 7 the recombinant antigens rCFP7A, rCFP17, rCFP21, rCFP25, and rCFP29 were all recognized in at least two strains of mice at a level comparable to ST-CF. rMPT51 and rCFP7 were only recognized in one or two strains respectively, at a level corresponding to no more than 1/3 of the response detected after ST-CF stimulation. Neither of the antigens rCFP20 and rCFP22 were recognized by any of the four mouse strains.
As shown in TABLE 7A, the recombinant antigens rCFP27, RD1-ORF2, rCFP10A, rCFP19, and rCFP25A were all recognized in at least two strains of mice at a level comparable to ST-CF, whereas rCFP23, and rCFP30B only were recognized in one strain at this level. rCFP30A , RD1-ORF5, rCFP16 gave rise to an IFN-γ release in two mice strains at a level corresponding to 2/3 of the response after stimulation with ST-CF. RD1-ORF3 was recognized in two strains at a level of 1/3 of ST-CF.

The native CFP7B was recognized in two strains at a level of 1/3 of the response seen after stimulation with ST-CF.

Memory responses. 8-12 weeks old female C57BL/6j(H-2^b) mice (Bomholtegaard, Ry) were given intravenous infections via the lateral tail vein with an inoculum of 5 x 10^4 *M. tuberculosis* suspended in PBS in a vol. of 0.1 ml. After 1 month of infection the mice were treated with isoniazid (Merck and Co., Rahway, NJ) and rifabutin (Farmatalia Carlo Erba, Milano, Italy) in the drinking water, for two months. The mice were rested for 4-6 months before being used in experiments. For the study of the recall of memory immunity, animals were infected with an inoculum of 1 x 10^6 bacteria i.v. and sacrificed at day 4 postinfection. Spleen cells were isolated and tested for the recognition of recombinant antigen.

As seen from TABLE 8, IFN-γ release after stimulation with rCFP17, rCFP21 and rCFP25 was at the same level as seen from spleen cells stimulated with ST-CF. Stimulation with rCFP7, rCFP7A and rCFP29 all resulted in an IFN-γ no higher than 1/3 of the response seen with ST-CF. rCFP22 was not recognized by IFN-γ producing cells. None of the antigens stimulated IFN-γ release in naive mice. Additionally non of the antigens were toxic to the cell cultures.

As shown in TABLE 8A, IFN-γ release after stimulation with RD1-ORF2 and rCFP19 was at the same level as seen from spleen cells stimulated with ST-CF. Stimulation with rCFP10A and rCFP30A gave rise to an IFN-γ release of 2/3 of the response after stimulation with ST-CF, whereas rCFP27, RD1-ORF5, rCFP23, rCFP25A and rCFP30B all resulted in an IFN-γ release no higher than 1/3 of the response seen with ST-CF. RD1-ORF3 and rCFP16 were not recognized by IFN-γ producing memory cells.

TABLE 7. T cell responses in primary TB infection.

Name	c57BL/6J($H2^b$)	DBA.2($H2^d$)	CBA/J($H2^k$)	A.SW($H2^s$)
rCFP7	+	+	-	-
rCFP7A	+++	+++	+++	+
rCFP17	+++	+	+++	+
rCFP20	-	-	-	-
rCFP21	+++	+++	+++	+
rCFP22	-	-	-	-
rCFP25	+++	++	+++	+
rCFP29	+++	+++	+++	++
rMPT51	+	-	-	-

Mouse IFN-γ release 14 days after primary infection with *M. tuberculosis.*

-:no response; +: 1/3 of ST-CF; ++: 2/3 of ST-CF; +++: level of ST-CF.

TABLE 7A. T cell responses in primary TB infection.

Name	C57Bl/6j ($H2^b$)	DBA.2 ($H2^d$)	CBA/J ($H2^k$)	A.SW ($H2^s$)
rCFP27	++	++	+++	+++
rCFP30A	-	+	++	++
RD1-ORF2	+++	+++	+++	++
RD1-ORF3	-	-	+	+
RD1-ORF5	+	+	++	++
rCFP10A	+++	n.d.	+++	n.d.
rCFP16	++	n.d.	++	n.d.
rCFP19	+++	n.d.	+++	n.d.
rCFP23	++	n.d.	+++	n.d.
rCFP25A	+++	n.d.	+++	n.d.
rCFP30B	+	n.d.	+++	n.d.
CFP7B(native)	+	n.d.	+	n.d.

Mouse IFN-γ release 14 days after primary infection with *M. tuberculosis.*

-: no response; +: 1/3 of ST-CF; ++: 2/3 of ST-CF; +++: level of ST-CF.

n.d. = not determined.

TABLE 8. T cell responses in memory immune animals.

Name	Memory response
rCFP7	+
rCFP7A	++
rCFP17	+++
rCFP21	+++
rCFP22	-
rCFP29	+
rCFP25	+++
rMPT51	+

Mouse IFN-γ release during recall of memory immunity to *M. tuberculosis.*

-:no response; +: 1/3 of ST-CF; ++: 2/3 of ST-CF; +++: level of ST-CF.

TABLE 8A. T cell responses in memory immune animals.

Name	Memory response
rCFP27	+
rCFP30A	++
RD1-ORF2	+++
RD1-ORF3	-
RD1-ORF5	+
rCFP10A	++
rCFP16	-
rCFP19	+++
rCFP23	+
rCFP25A	+
rCFP30B	+

Mouse IFN-γ release during recall of memory immunity to *M. tuberculosis.*

-: no response; +: 1/3 of ST-CF; ++: 2/3 of ST-CF; +++: level of ST-CF.

Interferon-γ induction in human TB patients and BCG vaccinated people.

Human donors: PBMC were obtained from healthy BCG vaccinated donors with no known exposure to patients with TB and from patients with culture or microscopy proven infection with *Mycobacterium tuberculosis*. Blood samples were drawn from the TB patients 1-4 months after diagnosis.

Lymphocyte preparations and cell culture: PBMC were freshly isolated by gradient centrifugation of heparinized blood on Lymphoprep (Nycomed, Oslo, Norway). The cells were resuspended in complete medium: RPMI 1640 (Gibco, Grand Island, N.Y.) supplemented with 40 µg/ml streptomycin, 40 U/ml penicillin, and 0.04 mM/ml glutamine, (all from Gibco Laboratories, Paisley, Scotland) and 10% normal human ABO serum (NHS) from the local blood bank. The number and the viability of the cells were determined by trypan blue staining. Cultures were established with 2,5 x 10^5 PBMC in 200 µl in microtitre plates (Nunc, Roskilde, Denmark) and stimulated with no antigen, ST-CF, PPD (2.5µg/ml); rCFP7, rCFP7A, rCFP17, rCFP20, rCFP21, rCFP22, rCFP25, rCFP26, rCFP29, in a final concentration of 5 µg/ml. Phytohaemagglutinin, 1 µg/ml (PHA, Difco laboratories, Detroit, MI. was used as a positive control. Supernatants for the detection of cytokines were harvested after 5 days of culture, pooled and stored at -80°C until use.

Cytokine analysis: Interferon-γ (IFN-γ) was measured with a standard ELISA technique using a commercially available pair of mAb´s from Endogen and used according to the instructions for use. Recombinant IFN-γ (Gibco laboratories) was used as a standard. The detection level for the assay was 50 pg/ml. The variation between the duplicate wells did not exceed 10 % of the mean. Responses of 9 individual donors are shown in TABLE 9.

A seen in TABLE 9 high levels of IFN-γ release are obtained after stimulation with several of the recombinant antigens. rCFP7a and rCFP17 gives rise to responses comparable to STCF in almost all donors. rCFP7 seems to be most strongly recognized by BCG vaccinated healthy donors. rCFP21, rCFP25, rCFP26, and rCFP29 gives rise to a mixed picture with intermediate responses in each group, whereas low responses are obtained by rCFP20 and rCFP22.

As is seen from Table 9A RD1-ORF2 and RD1-ORF5 give rise to IFN-γ responses close to the level of ST-CF. Between 60% and 90% of the donors show high IFN-γ responses (>1000 pg/ml).

rCFP30A gives rise to a mixed response with 40-50% high responders, whilst low responses are obtained with RD1-ORF3.

TABLE 9. Results from the stimulation of human blood cells from 5 healthy BCG vaccinated and 4 Tb patients with recombinant antigen. ST-CF, PPD and PHA are shown for comparison. Results are given in pg IFN-γ/ml.

Controls, Healthy, BCG vaccinated, no known TB exposure

donor:	no ag	PHA	PPD	STCF	CFP7	CFP17	CFP7A	CFP20	CFP21	CFP22	CFP25
1	6	9564	6774	3966	7034	69	1799	58	152	73	182
2	48	12486	6603	8067	3146	10044	5267	29	6149	51	1937
3	190	11929	10000	8299	8015	11563	8641	437	3194	669	2531
4	10	21029	4106	3537	1323	1939	5211	1	284	1	1344
5	1	18750	14209	13027	17725	8038	19002	1	3008	1	2103

TB patients, 1-4 month after diagnosis

	no ag	PHA	PPD	STCF	CFP7	CFP17	CFP7A	CFP20	CFP21	CFP22	CFP25
6	9	8973	5096	6145	852	4250	4019	284	1131	48	2400
7	1	12413	6281	3393	168	6375	4505	11	4335	16	3082
8	4	11915	7671	7375	104	2753	3356	119	407	437	2069
9	32	22130	16417	17213	8450	9783	16319	91	5957	67	10043

Table 9A.

Results from the stimulation of human blood cells from 10 healthy BCG vaccinated or non vaccinated ST-CF responsive healthy donors and 10 Tb patients with recombinant antigen are shown. ST-CF, PPD and PHA are included for comparison. Results are given in pg IFN-γ/ml and negative values below 300 pg/ml are shown as "<". nd = not done.

Controls, Healthy BCG vaccinated, or non vaccinated ST-CF positive

Donor	no ag	PHA	PPD	STCF	RD1-ORF2	RD1-ORF3	RD1-ORF5	rCFP30A
10	<	nd	3500	4200	1250	<	690	nd
11	<	nd	5890	4040	5650	880	9030	nd
12	<	nd	6480	3330	2310	<	3320	nd
13	<	nd	7440	4570	920	<	1230	nd
14	<	8310	nd	2990	1870	<	4880	<
15	<	10820	nd	4160	5690	<	810	3380
16	<	8710	nd	5690	1630	<	5600	<
17	<	7020	4480	5340	2030	nd	670	<
18	<	8370	6250	4780	3850	nd	370	1730
19	<	8520	1600	310	5110	nd	2330	1800

Tb patients, 1-4 month after diagnosis

Donor	no ag	PHA	PPD	STCF	RD1-ORF2	RD1-ORF3	RD1-ORF5	rCFP30A
20	<	nd	10670	12680	2020	<	9670	nd
21	<	nd	3010	1420	850	<	350	nd
22	<	nd	8450	7850	430	<	1950	nd
23	<	10060	nd	3730	<	<	350	<
24	<	10830	nd	6180	2090	<	320	730
25	<	9000	nd	3200	4760	<	4960	2820

26	<	10740	nd	7650	4710	<	1170	2280
27	<	7550	6430	6220	2030	nd	3390	3069
28	<	8090	5790	4850	1100	nd	2095	550
29	<	7790	4800	4260	2800	nd	1210	420

EXAMPLE 11

The recombinant antigens were tested individually as subunit vaccines in mice. Eleven groups of 6-8 weeks old, female C57Bl/6j mice (Bomholtegård, Denmark) were immunized subcutaneously at the base of the tail with vaccines of the following composition:

Group 1: 10 µg CFP7
Group 2: 10 µg CFP17
Group 3: 10 µg CFP21
Group 4: 10 µg CFP22
Group 5: 10 µg CFP25
Group 6: 10 µg CFP29
Group 7: 10 µg MPT51
Group 8: 50 µg ST-CF
Group 9: Adjuvant control group
Group 10: BCG 2,5 x 10^5/ml, 0,2 ml
Group 11: Control group: Untreated

All the subunit vaccines were given with DDA as adjuvant. The animals were vaccinated with a volume of 0.2 ml. Two weeks after the first injection and three weeks after the second injection group 1-9 were boosted a little further up the back. One week after the last injection the mice were bled and the blood cells were isolated. The immune response induced was

monitored by release of IFN-γ into the culture supernatant when stimulated in vitro with the homologous protein.

6 weeks after the last immunization the mice were aerosol challenged with 5 x 10^6 viable *Mycobacterium tuberculosis*/ml. After 6 weeks of infection the mice were killed and the number of viable bacteria in lung and spleen of infected mice was determined by plating serial 3-fold dilutions of organ homogenates on 7H11 plates. Colonies were counted after 2-3 weeks of incubation. The protective efficacy is expressed as the difference between $1og_{10}$ values of the geometric mean of counts obtained from five mice of the relevant group and the geometric mean of counts obtained from five mouse of the relevant control group.

The results from the experiments are presented in the following table.

Immunogenicity and protective efficacy in mice, of ST-CF and 7 subunit vaccines

Subunit Vaccine	Immunogenicity	Protective efficacy
ST-CF	+++	+++
CFP7	++	-
CFP17	+++	+++
CFP21	+++	++
CFP22	-	-
CFP25	+++	+++
CFP29	+++	+++
MPT51	+++	++

+++ Strong immunogen / high protection (level of BCG)

++ Medium immunogen / medium protection

- No recognition / no protection

In conclusion, we have identified a number of proteins inducing high levels of protection. Three of these CFP17, CFP25 and CFP29 giving rise to similar levels of protection as ST-CF

and BCG while two proteins CFP21 and MPT51 induces protections around 2/3 the level of BCG and ST-CF. Two of the proteins CFP7 and CFP22 did not induce protection in the mouse model.

As is described for rCFP7, rCFP17, rCFP21, rCFP22, rCFP25, rCFP29 and rMPT51 the two antigens rCFP7A and rCFP30A were tested individually as subunit vaccines in mice. C57Bl/6j mice were immunized as described for rCFP7, rCFP17, rCFP21, rCFP22, rCFP25, rCFP29 and rMPT51 using either 10μg rCFP7A or 10μg rCFP30A. Controls were the same as in the experiment including rCFP7, rCFP17, rCFP21, rCFP22, rCFP25, rCFP29 and rMPT51.

Immunogenicity and protective efficacy in mice, of ST-CF and 2 subunit vaccines.

Subunit vaccine	Immunogenicity	Protective efficacy
ST-CF	+++	+++
rCFP7A	+++	+++
rCFP30A	+++	-

+++	Strong immunogen/high protection (level of BCG)
++	Medium immunogen/medium protection
-	No recognition/no protection

In conclusion we have identified two strong immunogens of which one, rCFP7A, induces protection at the level of ST-CF.

EXAMPLE 12

Species distribution of cfp7, cfp9, mpt51, rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a and rd1-orf9b as well as of cfp7a, cfp7b, cfp10a, cfp17, cfp20, cfp21, cfp22, cfp22a, cfp23, cfp25 and *cfp25a.*

Presence of *cfp7, cfp9, mpt51, rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* in different mycobacterial species.

In order to determine the distribution of the *cfp7, cfp9,* mpt51*, rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* genes in species belonging to the *M. tuberculosis*-complex and in other mycobacteria PCR and/or Southern blotting was used. The bacterial strains used are listed in TABLE 10. Genomic DNA was prepared from mycobacterial cells as described previously (Andersen et al. 1992).

PCR analyses were used in order to determine the distribution of the *cfp7*, *cfp9* and *mpt51* gene in species belonging to the tuberculosis-complex and in other mycobacteria. The bacterial strains used are listed in TABLE 10. PCR was performed on genomic DNA prepared from mycobacterial cells as described previously (Andersen *et al.*, 1992).

The oligonucleotide primers used were synthesised automatically on a DNA synthesizer (Applied Biosystems, Forster City, Ca, ABI-391, PCR-mode), deblocked, and purified by ethanol precipitation. The primers used for the analyses are shown in TABLE 11.

The PCR amplification was carried out in a thermal reactor (Rapid cycler, Idaho Technology, Idaho) by mixing 20 ng chromosomal with the mastermix (contained 0.5 µM of each oligonucleotide primer, 0.25 µM BSA (Stratagene), low salt buffer (20 mM Tris-HCl, pH8.8 , 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$ and 0,1% Triton X-100) (Stratagene), 0.25 mM of each

deoxynucleoside triphosphate and 0.5 U Taq Plus Long DNA polymerase (Stratagene)). Final volume was 10 µl (all concentrations given are concentrations in the final volume). Predenaturation was carried out at 94°C for 30 s. 30 cycles of the following was performed: Denaturation at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 1 min.

The following primer combinations were used (the length of the amplified products are given in parentheses):

mpt51: MPT51-3 and MPT51-2 (820 bp), MPT51-3 and MPT51-6 (108 bp), MPT51-5 and MPT51-4 (415 bp), MPT51-7 and MPT51-4 (325 bp).
cfp7: pVF1 and PVR1 (274 bp), pVF1 and PVR2 (197 bp), pVF3 and PVR1 (302 bp), pVF3 and PVR2 (125 bp).
cfp9: stR3 and stF1 (351 bp).

TABLE 10.
Mycobacterial strains used in this Example.

Species and strain(s)		Source
1. *M. tuberculosis* (ATCC 27294)	H37Rv	ATCC[a]
2. (ATCC 25177)	H37Ra	ATCC
3.	Erdman	Obtained from A. Lazlo, Ottawa, Canada
4. *M. bovis* BCG substrain:	Danish 1331	SSI[b]
5.	Chinese	SSI[c]
6.	Canadian	SSI[c]
7.	Glaxo	SSI[c]
8.	Russia	SSI[c]
9.	Pasteur	SSI[c]
10.	Japan	WHO[e]
11. *M. bovis* MNC 27		SSI[c]
12. *M. africanum*		Isolated from a Danish patient
13. *M. leprae* (armadillo-derived)		Obtained from J. M. Colston, London, UK
14. *M. avium* (ATCC 15769)		ATCC
15. *M. kansasii* (ATCC 12478)		ATCC
16. *M. marinum* (ATCC 927)		ATCC
17. *M. scrofulaceum* (ATCC 19275)		ATCC
18. *M. intercellulare* (ATCC 15985)		ATCC
19. *M. fortuitum* (ATCC 6841)		ATCC
20. *M. xenopi*		Isolated from a Danish patient
21. *M. flavescens*		Isolated from a Danish patient
22. *M. szulgai*		Isolated from a Danish patient
23. *M. terrae*		SSI[c]
24. *E. coli*		SSI[d]
25. *S.aureus*		SSI[d]

[a] American Type Culture Collection, USA.

[b] Statens Serum Institut, Copenhagen, Denmark.

[c] Our collection Department of Mycobacteriology, Statens Serum Institut, Copenhagen, Denmark.

[d] Department of Clinical Microbiology, Statens Serum Institut, Denmark.

[e] WHO International Laboratory for Biological Standards, Statens Serum Institut, Copenhagen, Denmark.

TABLE 11. Sequences of the *mpt51*, *cfp7* and *cfp9* oligonucleotides.

Orientation and oligonucleotide	Sequences (5'→3')[a]	Position[b] (nucleotides)
Sense		
MPT51-1	CTCGAATTCGCCGGGTGCACACAG (SEQ ID NO: 28)	6 - 21 (SEQ ID NO: 41)
MPT51-3	CTCGAATTCGCCCCATACGAGAAC (SEQ ID NO: 29)	143 - 158 (SEQ ID NO: 41)
MPT51-5	GTGTATCTGCTGGAC (SEQ ID NO: 30)	228 - 242 (SEQ ID NO: 41)
MPT51-7	CCGACTGGCTGGCCG (SEQ ID NO: 31)	418 - 432 (SEQ ID NO: 41)
pvR1	GTACGAGAATTCATGTCGCAAATCATG (SEQ ID NO: 35)	91 - 105 (SEQ ID NO: 1)
pvR2	GTACGAGAATTCGAGCTTGGGGTGCCG (SEQ ID NO: 36)	168 - 181 (SEQ ID NO: 1)
stR3	CGATTCCAAGCTTGTGGCCGCCGACCCG (SEQ ID NO: 37)	141 - 155 (SEQ ID NO: 3)
Antisense		
MPT51-2	GAGGAATTCGCTTAGCGGATCGCA (SEQ ID NO: 32)	946 - 932 (SEQ ID NO: 41)
MPT51-4	CCCACATTCCGTTGG (SEQ ID NO: 33)	642 - 628 (SEQ ID NO: 41)
MPT51-6	GTCCAGCAGATACAC (SEQ ID NO: 34)	242 - 228 (SEQ ID NO: 41)
pvF1	CGTTAGGGATCCTCATCGCCATGGTGTTGG (SEQ ID NO: 38)	340 - 323 (SEQ ID NO: 1)
pvF3	CGTTAGGGATCCGGTTCCACTGTGCC (SEQ ID NO: 39)	268 - 255 (SEQ ID NO: 1)
stF1	CGTTAGGGATCCTCAGGTCTTTTCGATG (SEQ ID NO: 40)	467 - 452 (SEQ ID NO: 3)

[a] Nucleotides underlined are not contained in the nucleotide sequences of *mpt*51, *cfp*7, and *cfp*9.

[b] The positions referred to are of the non-underlined parts of the primers and correspond to the nucleotide sequence shown in SEQ ID NOs: 41, 1, and 3 for *mpt*51, *cfp*7, and *cfp*9, respectively.

The Southern blotting was carried out as described previously (Oettinger and Andersen, 1994) with the following modifications: 2 μg of genomic DNA was digested with *Pvu*II, electrophoresed in an 0.8% agarose gel, and transferred onto a nylon membrane (Hybond N-plus; Amersham International plc, Little Chalfont, United Kingdom) with a vacuum transfer device (Milliblot, TM-v; Millipore Corp., Bedford, MA). The *cfp7, cfp9,* mpt51, *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* gene fragments were amplified by PCR from the plasmids pRVN01, pRVN02, pTO52, pTO87, pTO88, pT089, pT090, pT091, pT096 or pTO98 by using the primers shown in TABLE 11 and TABLE 2. The probes were labelled non-radioactively with an enhanced chemiluminescence kit (ECL; Amersham International plc, Little Chalfont, United Kingdom). Hybridization and detection was performed according to the instructions provided by the manufacturer. The results are summarized in TABLES 12 and 13.

TABLE 12. Interspecies analysis of the *cfp7, cfp9* and *mpt51* genes by PCR and/or Southern blotting and of MPT51 protein by Western blotting.

		PCR			Southern blot			Western blot
Species and strain		*cfp7*	*cfp9*	mpt51	*cfp7*	*cfp9*	mpt51	MPT51
1.	*M. tub.* H37Rv	+	+	+	+	+	+	+
2.	*M. tub.* H37Ra	+	+	+	N.D.	N.D.	+	+
3.	*M. tub.* Erdmann	+	+	+	+	+	+	+
4.	*M. bovis*	+	+	+			+	+
5.	*M. bovis* BCG Danish 1331	+	+	+	+	+	+	+
6.	*M. bovis* BCG Japan	+	+	N.D.	+	+	+	N.D.
7.	*M. bovis* BCG Chinese	+	+	N.D.	+	+	N.D.	N.D.
8.	*M. bovis* BCG Canadian	+	+	N.D.	+	+	N.D.	N.D.
9.	*M. bovis* BCG Glaxo	+	+	N.D.	+	+	N.D.	N.D.
10.	*M. bovis* BCG Russia	+	+	N.D.	+	+	N.D.	N.D.
11.	*M. bovis* BCG Pasteur	+	+	N.D.	+	+	N.D.	N.D.
12.	*M. africanum*	+	+	+	+	+	+	+
13.	*M. leprae*	-	-	-	-	-	-	-
14.	*M. avium*	+	+	-	+	+	+	-
15.	*M. kansasii*	+	-	-	+	+	+	-
16.	*M. marinum*	-	(+)	-	+	+	+	-
17.	*M. scrofulaceum*	-	-	-	-	-	-	-
18.	*M. intercellulare*	+	(+)	-	+	+	+	-
19.	*M. fortuitum*	-	-	-	-	-	-	-
20.	*M. flavescens*	+	(+)	-	+	+	+	N.D.
21.	*M. xenopi*	-	-	-	N.D.	N.D.	+	-
22.	*M. szulgai*	(+)	(+)	-	-	+	-	-
23.	*M. terrae*	-	-	N.D.	N.D.	N.D.	N.D.	N.D.

+, positive reaction; -, no reaction, N.D. not determined.

cfp7, cfp9 and *mpt51* were found in the *M. tuberculosis* complex including BCG and the environmental mycobacteria; *M. avium, M. kansasii, M. marinum, M. intracellular* and *M. flavescens. cfp9* was additionally found in *M. szulgai* and *mpt51* in *M. xenopi.*

Furthermore the presence of native MPT51 in culture filtrates from different mycobacterial strains was investigated with western blots developed with Mab HBT4.

There is a strong band at around 26 kDa in *M. tuberculosis* H37Rv, Ra, Erdman, *M. bovis* AN5, *M. bovis* BCG substrain Danish 1331 and *M. africanum.* No band was seen in the region in any other tested mycobacterial strains.

TABLE 13a. Interspecies analysis of the *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* genes by Southern blotting.

Species and strain	*rd1-orf2*	*rd1-orf3*	*rd1-orf4*	*rd1-orf5*	*rd1-orf8*	*rd1-orf9a*	*rd1-orf9b*
1. M. tub. H37Rv	+	+	+	+	+	+	+
2. M. bovis	+	+	+	+	N.D.	+	+
3. M. bovis BCG Danish 1331	+	-	-	-	N.D.	-	-
4. M. bovis BCG Japan	+	-	-	-	N.D.	-	-
5. M. avium	-	-	-	-	N.D.	-	-
6. M. kansasii	-	-	-	-	N.D.	-	-
7. M. marinum	+	-	+	-	N.D.	-	-
8. M. scrofulaceum	+	-	-	-	N.D.	-	-
9. M. intercellulare	-	-	-	-	N.D.	-	-
10. M. fortuitum	-	-	-	-	N.D.	-	-
11. M. xenopi	-	-	-	-	N.D.	-	-
12. M. szulgai	+	-	-	-	N.D.	-	-

+, positive reaction; -, no reaction, N.D. not determined.

Positive results for *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* were only obtained when using genomic DNA from *M. tuberculosis* and *M. bovis*, and not from *M. bovis* BCG or other mycobacteria analyzed except *rd1-orf4* which also was found in *M. marinum*.

Presence of *cfp7a, cfp7b, cfp10a, cfp17, cfp20, cfp21, cfp22, cfp22a, cfp23, cfp25* and *cfp25a* in different mycobacterial species.

Southern blotting was carried out as described for *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b*. The *cfp7a, cfp7b, cfp10a, cfp17, cfp20, cfp21, cfp22, cfp22a, cfp23, cfp25* and *cfp25a* gene fragments were amplified by PCR from the recombinant pMCT6 plasmids encoding the individual genes. The primers used (same as the primers used for cloning) are described in example 3, 3A and 3B. The results are summarized in Table 13b.

TABLE 13b. Interspecies analysis of the *cfp7a, cfp7b, cfp10a, cfp17, cfp20, cfp21, cfp22, cfp22a, cfp23, cfp25,* and *cfp25a* genes by Southern blotting.

Species and strain	*cfp7a*	*cfp7b*	*cfp-10a*	*cfp17*	*cfp20*	*cfp21*	*cfp22*	*cfp-22a*	*cfp23*	*cfp25*	*cfp-25a*
1. M. tub. H37Rv	+	+	+	+	+	+	+	+	+	+	+
2. M. bovis	+	+	+	+	+	+	+	+	+	+	+
3. M. bovis BCG Danish 1331	+	+	+	+	+	N.D.	+	+	+	+	+
4. M. bovis BCG Japan	+	+	+	+	+	+	+	+	+	+	+
5. M. avium	+	N.D.	-	+	-	+	+	+	+	+	-
6. M. kansasii	-	N.D.	+	-	-	-	+	-	+	-	-
7. M. marinum	+	+	-	+	+	+	+	+	+	+	+
8. M. scrofulaceum	-	-	+	-	+	+	-	+	+	+	-
9. M. intercellulare	+	+	-	+	-	+	+	-	+	+	-
10. M. fortuitum	-	N.D.	-	-	-	-	-	-	+	-	-
11. M. xenopi	+	+	+	+	+	+	+	+	+	+	+
12. M. szulgai	+	+	-	+	+	+	+	+	+	+	+

+, positive reaction; -, no reaction, N.D. not determined.

EXAMPLE 13:

Identification of the immunogenic portions of the three antigenic molecules TB10.3 (Rv3019c), TB10.4 (Rv0288) and TB12.9 (Rv3017c).

The three immunologically active proteins, of which we are here identifying the immunogenic portions, were previously identified by the screening of a genomic library (TB10.4, - previously named CFP7) and due to homology to TB10.4 (TB10.3 and TB12.9) (WO98/44119, WO99/24577 and Skjøt et al, 2000). However, the immunogenic portions of these proteins have not previously been defined.

Synthetic overlapping peptides covering the complete amino acid sequence of the three proteins (figures 13, 14 and 15) were synthesized either by standard solid-phase methods on an ABIMED peptide synthesizier (ABIMED, Langenfeld, Germany) at Dept. of infectious diseases and Immunohematology/Bloodbank C5-P, Leiden University Medical Centre, The Netherlands (TB10.4) or at Schafer-N, Copenhagen, Denmark on polyamide resins using Fmoc-strategy and purified by reverse phase HPLC on C18-columns in water/acetonitrile gradients containing 0.1%TFA (trifluoracetic acid) (TB10.3 and TB12.9).
Purified peptides were lyophilized and stored dry until reconstitution in PBS.

The peptide sequences are as follows;

TB10.3:

TB10.3-P1 MSQIMYNYPAMMAHAGDMAG
TB10.3-P2 MMAHAGDMAGYAGTLQSLGA
TB10.3-P3 YAGTLQSLGADIASEQAVLS
TB10.3-P4 DIASEQAVLSSAWQGDTGIT
TB10.3-P5 SAWQGDTGITYQGWQTQWNQ
TB10.3-P6 YQGWQTQWNQALEDLVRAYQ
TB10.3-P7 ALEDLVRAYQSMSGTHESNT

TB10.3-P8 SMSGTHESNTMAMLARDGAE

TB10.3-P9 MAMLARDGAEAAKWGG

TB10.4:

TB10.4-P1 MSQIMYNYPAMLGHAGDM

TB10.4-P2 MLGHAGDMAGYAGTLQSL

TB10.4-P3 YAGTLQSLGAEIAVEQAA

TB10.4-P4 EIAVEQAALQSAWQGDTG

TB10.4-P5 SAWQGDTGITYQAWQAQW

TB10.4-P6 YQAWQAQWNQAMEDLVRA

TB10.4-P7 AMEDLVRAYHAMSSTHEA

TB10.4-P8 AMSSTHEANTMAMMARDT

TB10.4-P9 MAMMARDTAEAAKWGG

TB12.9:

TB12.9-P1 MSQSMYSYPAMTANVGDMAG

TB12.9-P2 MTANVGDMAGYTGTTQSLGA

TB12.9-P3 YTGTTQSLGADIASERTAPS

TB12.9-P4 DIASERTAPSRACQGDLGMS

TB12.9-P5 RACQGDLGMSHQDWQAQWNQ

TB12.9-P6 HQDWQAQWNQAMEALARAYR

TB12.9-P7 AMEALARAYRRCRRALRQIG

TB12.9-P8 RCRRALRQIGVLERPVGDSS

TB12.9-P9 VLERPVGDSSDCGTIRVGSF

TB12.9-P10 DCGTIRVGSFRGRWLDPRHA

TB12.9-P11 RGRWLDPRHAGPATAADAGD

Peptides are encoded by the following DNA sequences:

tb10.3:

tb10.3-P1 atg tcg cag att atg tac aac tat ccg gcg atg atg gct cat gcc ggg gac atg gcc ggt

tb10.3-P2 atg atg gct cat gcc ggg gac atg gcc ggt tat gcg ggc acg ctg cag agc ttg ggg gcc

tb10.3-P3 tat gcg ggc acg ctg cag agc ttg ggg gcc gat atc gcc agt gag cag gcc gtg ctg tcc

tb10.3-P4 gat atc gcc agt gag cag gcc gtg ctg tcc agt gct tgg cag ggt gat acc ggg atc acg

tb10.3-P5 agt gct tgg cag ggt gat acc ggg atc acg tat cag ggc tgg cag acc cag tgg aac cag

tb10.3-P6 tat cag ggc tgg cag acc cag tgg aac cag gcc cta gag gat ctg gtg cgg gcc tat cag

tb10.3-P7 gcc cta gag gat ctg gtg cgg gcc tat cag tcg atg tct ggc acc cat gag tcc aac acc

tb10.3-P8 tcg atg tct ggc acc cat gag tcc aac acc atg gcg atg ttg gct cga gat ggg gcc gaa

tb10.3-P9 atg gcg atg ttg gct cga gat ggg gcc gaa gcc gcc aag tgg ggc ggc

tb10.4:

TB10.4-P1 atg tcg caa atc atg tac aac tac ccc gcg atg ttg ggt cac gcc ggg gat atg

TB10.4-P2 atg ttg ggt cac gcc ggg gat atg gcc gga tat gcc ggc acg ctg cag agc ttg

TB10.4-P3 tat gcc ggc acg ctg cag agc ttg ggt gcc gag atc gcc gtg gag cag gcc gcg

TB10.4-P4 gag atc gcc gtg gag cag gcc gcg ttg cag agt gcg tgg cag ggc gat acc ggg

TB10.4-P5 agt gcg tgg cag ggc gat acc ggg atc acg tat cag gcg tgg cag gca cag tgg

TB10.4-P6 tat cag gcg tgg cag gca cag tgg aac cag gcc atg gaa gat ttg gtg cgg

TB10.4-P7 gcc atg gaa gat ttg gtg cgg gcc tat cat gcg atg tcc agc acc cat gaa gcc

TB10.4-P8 gcg atg tcc agc acc cat gaa gcc aac acc atg gcg atg atg gcc cgc gac acg

TB10.4-P9 atg gcg atg atg gcc cgc gac acc gcc gaa gcc gcc aaa tgg ggc ggc

tb12.9:

tb12.9-P1 gtg tcg cag agt atg tac agc tac ccg gcg atg acg gcc aat gtc gga gac atg gcc ggt

tb12.9-P2 atg acg gcc aat gtc gga gac atg gcc ggt tat acg ggc acg acg cag agc ttg ggg gcc

tb12.9-P3 tat acg ggc acg acg cag agc ttg ggg gcc gat atc gcc agt gag cgc acc gcg ccg tcg

tb12.9-P4 gat atc gcc agt gag cgc acc gcg ccg tcg cgt gct tgc caa ggt gat ctc ggg atg agt

tb12.9-P5 cgt gct tgc caa ggt gat ctc ggg atg agt cat cag gac tgg cag gcc cag tgg aat cag

tb12.9-P6 cat cag gac tgg cag gcc cag tgg aat cag gcc atg gag gct ctc gcg cgg gcc tac cgt

tb12.9-P7 gcc atg gag gct ctc gcg cgg gcc tac cgt cgg tgc cgg cga gca cta cgc cag atc ggg

tb12.9-P8 cgg tgc cgg cga gca cta cgc cag atc ggg gtg ctg gaa agg ccg gta ggc gat tcg tca
tb12.9-P9 gtg ctg gaa agg ccg gta ggc gat tcg tca gac tgc gga acg att agg gtg ggg tcg ttc
tb12.9-P10 gac tgc gga acg att agg gtg ggg tcg ttc cgg ggt cgg tgg ctg gac ccg cgc cat gcg
tb12.9-P11 cgg ggt cgg tgg ctg gac ccg cgc cat gcg ggt cca gcc acg gcc gcc gac gcc gga gac

In tb12.9-P1, the start codon is indicated to be gtg, and the first amino acid in TB 12.9-P1 is indicated to be methionin. However, it is supposed that the first amino acid in TB 12.9 can be either methionin or valin. The invention therefore encompasses TB12.9-P1 and TB12.9 polypeptide sequences with amino acid sequences as indicated in this description and drawing, and sequences in which the first amino acid is replaced with valin. Accordingly, the invention encompasses tb12.9-P1 and tb12.9 DNA sequences as indicated in this description and drawing, and DNA sequences in which the gtg start codon is replaced with an atg start codon.

EXAMPLE 14:

Biological activity of the synthetic peptides.

The above listed synthetic peptides, covering the protein sequences of TB10.3, TB10.4 and TB12.9, were screened for their ability to induce a T cell response measured as IFN-γ release. The screening involved testing of the IFN-γ induction in PBMC preparations obtained from TB patients and PPD positive donors.

Human donors: PBMC were obtained from healthy donors with a positive *in vitro* response to purified protein derivative (PPD) or non-vaccinated healthy donors with a negative *in vitro* response to PPD. PBMC were also obtained from TB patients with microscopy or culture proven infection. Blood samples were drawn from TB patients 0-6 months after diagnosis.

Lymphocyte preparations and cell culture: PBMC were freshly isolated by gradient centrifugation of heparinized blood on Lymphoprep (Nycomed, Oslo, Norway) and stored in liquid nitrogen until use. The cells were resuspended in complete RPMI 1640 medium (Gibco BRL, Life Technologies) supplemented with 1% penicillin/streptomycin (Gibco BRL, Life Technologies), 1% non-essential-amino acids (FLOW, ICN Biomedicals, CA, USA), and 10% heat-inactivated normal human AB serum (NHS). The viability and number of the cells were determined by Nigrosin staining. Cell cultures were established with 1.25 x 10^5 PBMCs in 100 µl in microtitre plates (Nunc, Roskilde, Denmark) and stimulated with 5 µg/ml PPD or single peptides in final concentrations of 10, 5 and 0.5µg/ml or peptide pools corresponding to the full length proteins in which each peptide was included in concentrations of 0.5, 0.1 and 0.05µg/ml (Table 1). "No antigen" was included as negative control and phytohaemagglutinin (PHA) was used as positive control. Moreover the response to the highly responsive recombinant TB10.4 was included for comparison. Supernatants for the analysis of secreted cytokines were harvested after 5 days of culture, pooled, and stored at -80 °C until use.

Cytokine analysis:

In table 14 the maximal IFN-γ response of each peptide is shown. A response of > 75pg IFN-γ/ml (> background plus two standard deviations) is regarded positive (indicated in bold).
In PPD negative healthy donors, the IFN-γ response to the peptides was at the same level as the buffer control (results not shown).
As shown in Table 14 the peptides TB10.3-P1, TB10.3-P2, TB10.3-P3, TB10.3-P4, TB10.3-P5, TB10.3-P6, TB10.3-P7, TB10.3-P8, TB10.3-P9, TB10.4-P1, TB10.4-P2, TB10.4-P3, TB10.4-P4, TB10.4-P5, TB10.4-P6, TB10.4-P7, TB10.4-P8, TB10.4-P9, TB12.9-P1, TB12.9-P2, TB12.9-P4, TB12.9-P5, TB12.9-P6, TB12.9-P7, TB12.9-P8, TB12.9-P9, TB12.9-P10 and TB12.9-P11 result in positive responses. As is expected, due to the genetical heterogeneity

of the human population, some of the peptides are however recognized more frequently and to a higher extent than others.

The immunodominant peptides in the 8 donors tested herein are the peptides TB10.3-P1, TB10.4-P1, TB10.4-P3, TB10.4-P5, TB10.4-P6, and TB10.4-P9 which all give rise to IFN-γ responses >1000pg/ml in at least one of the 8 donors tested herein. Of these peptides TB10.4-p1 and TB10.4-p3 give rise to responses at the level of the full protein in several donors and two donors (C and D in table 14) respond to TB10.4-p1 at a level comparable to PPD.

Other peptides that are highly responsive are TB10.3-P3, TB10.3-P6, TB10.3-P8, TB10.4-P4, TB12.9-P1, TB12.9-P8, TB12.9-P9, and TB12.9-P11 which give rise to IFN-γ responses > 450pg/ml (< 1000pg/ml) in at least one of the 8 donors tested herein.

The peptides; TB10.3-P2, TB10.3-P9, TB12.9-P2, TB12.9-P4 and TB12.9-P10 induce lower but still highly significant levels of IFN-γ > 250pg/ml (<450pg/ml) in at least one of the 8 donors tested herein.

The surprisingly broad recognition pattern indicates the presence of multiple immunogenic portions scattered through out the protein sequences of TB10.3, TB10.4 and TB12.9. This makes these peptides attractive candidates for a TB vaccine or a therapeutic vaccine.

Table 14. Stimulation of PBMCs from 4 TB patients and 4 PPD positive healthy donors with synthetic peptides. Responses to PHA, PPD, no antigen and to rTB10.4 are shown for comparison. Results are given in pg IFN-γ /ml. TB10.4 p1-9, TB10.3 p1-9, TB12.9 p1-9 indicate peptide pools. Results indicated in bold are regarded positive.

Peptide/ Antigen	TB patients				PPD + healthy donors			
	1	2	3	4	A	B	C	D
No antigen	7	0	21	3	5	3	0	0
PHA	16493	12187	13217	13458	14941	13051	10801	9112
PPD	5579	10164	8746	6697	16797	8963	6768	828
rTB10.4	63	1766	610	3354	3295	2688	4296	1708
TB10.4 p1	48	**550**	**648**	**421**	0	**2149**	**7120**	**1007**
TB10.4 p2	**132**	**125**	**103**	**113**	51	2	3	0
TB10.4 p3	**306**	**1207**	7	**98**	**3121**	7	0	**638**
TB10.4 p4	**248**	7	**102**	**226**	**79**	8	0	**593**
TB10.4 p5	**136**	0	0	**1377**	57	**837**	0	0
TB10.4 p6	36	0	**2183**	**180**	24	13	0	0
TB10.4 p7	74	27	0	**164**	**184**	5	0	0
TB10.4 p8	67	27	**117**	46	**128**	7	0	0
TB10.4 p9	14	0	1	**1034**	59	**91**	13	0
TB10.4 p1-9	**685**	**4495**	**3056**	**1435**	**2029**	**2981**	**3424**	**605**
TB12.9 p1	16	0	**224**	**486**	**117**	0	0	0
TB12.9 p2	**121**	0	50	**278**	**125**	11	0	0
TB12.9 p3	14	1	33	48	16	21	0	0
TB12.9 p4	**225**	21	**269**	**149**	14	7	0	0
TB12.9 p5	0	91	**95**	**119**	32	4	0	0
TB12.9 p6	0	0	**78**	45	20	49	0	44
TB12.9 p7	0	64	1	61	63	**99**	0	15
TB12.9 p8	0	**95**	0	**977**	19	**966**	40	**125**
TB12.9 p9	0	0	**505**	67	15	2	0	29
TB12.9 p10	54	0	**206**	**369**	17	1	0	6
TB12.9 p11	54	0	**118**	**484**	13	0	0	9
TB12.9 p1-11	**256**	0	0	**204**	11	**217**	0	13
TB10.3 p1	**88**	**449**	30	**76**	**196**	**1358**	**706**	**122**
TB10.3 p2	**99**	**214**	**341**	**172**	48	**387**	20	13
TB10.3 p3	3	**408**	**204**	**460**	**275**	3	0	21

TB10.3 p4	0	41	66	50	**210**	0	0	26
TB10.3 p5	0	30	**185**	**215**	**184**	11	0	0
TB10.3 p6	0	**138**	**465**	21	22	5	0	2
TB10.3 p7	11	**143**	**160**	30	37	27	0	9
TB10.3 p8	**744**	3	4	**92**	15	**663**	19	0
TB10.3 p9	**341**	0	6	11	66	0	2	0
TB10.3 p1-9	0	**187**	**193**	3	8	**903**	**262**	6

References

Andersen P. *et al.*, 1995, J. Immunol. 154: 3359-72

Andersen P., 1994, Infect. Immun. 62: 2536-44.

Andersen, P., and Heron, I. 1993 J. Immunol. Methods 161 29-39

Andersen, P., D. Askgaard, L. Ljungqvist, J. Bennedsen, I. Heron. 1991a. Infection and Immunity 59: 1558-1563.

Andersen, P., D. Askgaard, L. Ljungqvist, J. Bennedsen, I. Heron. 1991b. Infection and Immunity. 59: 1905-1910.

Andersen, Å. B. *et al.*, 1992, Infect. Immun. 60: 2317-2323.

Barkholt, V. and Jensen, A. L., 1989, Anal. Biochem. 177: 318-322.

Borodovsky, M., and J. McIninch. 1993, Computers Chem. 17: 123-133.

Brandt, L., *et al.* 2000 Infect. Immun. 68:2; 791-795.

Cole, S.T *et al* 1998 Nature 393: 537-544

Cote-Sierra J, *et al* 1998, Gene Oct 9;221(1):25-34

Flesch, I., and S.H.E. Kaufmann. 1987. J. Immunol. 138: 4408-4413

Gosselin *et a*l 1992 J. Immunol. 149; 3477-3481

Harboe, M. *et al.*, 1996, Infect. Immun. 64: 16-22.

Harboe, M., *et al* 1998 Infect. Immun. 66:2; 717-723

Hochstrasser, D.F. *et al.*, 1988, Anal.Biochem. 173: 424-435

Kilgus J et al, J Immunol. 1991 Jan 1;146(1):307-15.

Köhler, G. and Milstein, C., 1975, Nature 256: 495-497.

Lefford, M.J., and D.D. McGregor. 1974. Cell. Immunol. 14: 417-428.

Li, H. *et al.*, 1993, Infect. Immun. 61: 1730-1734.

Lindblad E.B. *et al.*, 1997, Infect. Immun. 65: 623-629.

Lowry, D.B. *et al* 1999, Nature 400: 269-71

Mahairas, G. G. *et al.*, 1996, J. Bacteriol 178: 1274-1282.

Maniatis T. *et al.*, 1989, "Molecular cloning: a laboratory manual", 2nd ed., Cold Spring harbor Laboratory, Cold Spring Harbor, N.Y.

McCafferty et al., Nature, 348:552-554 (1990).

Nagai *et al* 1991, Infect. Immun 59:1; 372-382.

Oettinger, T. and Andersen, Å. B., 1994, Infect. Immun. 62: 2058-2064.

Ohara, N. *et al.*, 1995, Scand. J. immunol. 41: 233-442.

Olsen AW *et al* Eur J Immunol. 2000 Jun;30(6):1724-32.

Orme, I.M. 1988. J. Immunol. 140: 3589-3593.

Pal P. G. and Horwitz M. A., 1992, Infect. Immun. 60: 4781-92.

Patent application WO 01/04151 entitled "Tuberculosis vaccine and diagnostic based on the *Mycobacterium tuberculosis* esat-6 gene family".

Patent application PA 2000 00666 entitled " Nucleic acid fragments and polypeptide fragments derived from *M. tuberculosis*"

Patent application US 09/0505,739 "Nucleic acid fragments and polypeptide fragments derived from *M. tuberculosis*"

Pearson, W. R. and Lipman D. J., 1988. Proc. Natl. Acad. Sci. USA 85: 2444-2448.

Ploug, M. *et al.*, 1989, Anal. Biochem. 181: 33-39.

Pollock. J., et al, 2000. The Veterinary record, 146:659-665

Porath, J. *et al.*, 1985, FEBS Lett. 185: 306-310.

Ravn et al. 1999, Journal of infectious diseases 179:637-45

Roberts, A.D. *et al.*, 1995, Immunol. 85: 502-508.

Rolph, M.S, and I. A. Ramshaw. 1997. Curr.Opin.Immunol.9:517-24

Rook, G.A.W. 1990. Res. Microbiol. 141: 253-256.

Rosenkrands, I., *et al* 1998, Infect. Immun 66:6; 2728-2735

Sambrook *et al* (Molecular Cloning; A laboratory manual, Cold Spring Harbor Laboratories, NY, 1989

Sinigaglia F et al. Nature 1988 Dec 22-29;336(6201):778-80

Skjøt, R.L.V., *et al* 2000, Infect. Immun 68:1; 214-220

Stryhn, A., *et al* 1996 Eur. J. Immunol. 26:1911-1918

Sørensen, A.L. *et al.*, 1995, Infect. Immun. 63: 1710-1717.

Theisen, M. *et al.*, 1995, Clinical and Diagnostic Laboratory Immunology, 2: 30-34.

Thompson J., *et al* Nucleic Acids Res 1994 22:4673-4680

Ulmer J.B *et al* 1993, Curr. Opin. Invest. Drugs 2(9): 983-989

Valdés-Stauber, N. and Scherer, S., 1994, Appl. Environ. Microbiol. 60: 3809-3814.

Valdés-Stauber, N. and Scherer, S., 1996, Appl. Environ. Microbiol. 62: 1283-1286.

van Dyke M. W. *et al.*, 1992. Gene pp. 99-104.

von Heijne, G., 1984, J. Mol. Biol. 173: 243-251.

Williams, N., 1996, Science 272: 27.

Young, R. A. *et al.*, 1985, Proc. Natl. Acad. Sci. USA 82: 2583-2587.